CFTRI-M [illegible]
3871
Textbook of phys..

TEXTBOOK OF
PHYSIOLOGY AND
BIOCHEMISTRY

This book is protected under the Berne Convention. It may not be reproduced by any means, in whole or in part, without permission. Application with regard to reproduction should be addressed to the Publishers.

TEXTBOOK OF PHYSIOLOGY AND BIOCHEMISTRY

B

BY

GEORGE H. BELL
B.Sc., M.D.(Glasg.), F.R.F.P.S.G., F.R.S.E.
Professor of Physiology in the University of St Andrews at Queen's College, Dundee

J. NORMAN DAVIDSON
M.D., D.Sc.(Edin.), F.R.F.P.S.G., F.R.I.C., F.R.S.E.
Gardiner Professor of Physiological Chemistry in the University of Glasgow; formerly Professor of Biochemistry in the University of London at St Thomas's Hospital Medical School

AND

HAROLD SCARBOROUGH
M.B., Ph.D.(Edin.), F.R.C.P.E., M.R.C.P.
Professor of Medicine in the Welsh National School of Medicine of the University of Wales and Director of the Medical Unit in the Royal Infirmary, Cardiff; formerly Reader in Medicine in the University of Birmingham.

With a foreword by

ROBERT C. GARRY
M.B., D.Sc.(Glasg.), F.R.F.P.S.G., F.R.S.E.
Regius Professor of Physiology in the University of Glasgow

Third Edition Reprint

CENTRAL FOOD TECHNOLOGICAL RESEAR
30 JUN 1958
LIBRARY
* MYSORE. *

E. & S. LIVINGSTONE LTD
EDINBURGH AND LONDON
1957

First Edition . . 1950
Reprinted . . 1952
Second Edition . 1953
Reprinted . . 1954
Third Edition . . 1956
Reprinted . . 1957

3871 ✓

[redacted]

L;3&lGxt N56

CFTRI-MYSORE
3871
Textbook of phys..

Printed in Great Britain by Neill & Co. Ltd., Edinburgh.

FOREWORD

IN all sciences, as the body of knowledge grows, specialization becomes inevitable. A point is reached, often suddenly, when the interests of the various workers in a science diverge. Then each worker pursues his own line of enquiry, dropping, as he goes, a curtain of new techniques and of new knowledge between himself and his former colleagues. Physiology is peculiarly vulnerable to this process. The basic sciences, Physics and Chemistry, continually offer to the physiologist new weapons to attack biological problems. These weapons demand in their use special training and skill. Thus disintegration advances knowledge.

Nevertheless, all workers in Physiology are held together by the fact that the phenomena they study form part of a single whole, the living organism. However dissimilar the methods of approach, there is unity in the object attacked. An organism is an individual entity, a whole maintaining its integrity in a physical and chemical environment throughout a period of time. Thus the significance of the most recondite research cannot be fully appreciated without a knowledge of the entire living background which is the organism itself. Minor modification of a statement by Michael Foster gives a true picture of the position.

"We may speak of an organism as a complex structure, but we must strive to realize that what we mean by that is a complex whirl, an intricate dance, of which what we call biophysical activity, biochemical reactions, histological structure and gross configuration are, so to speak, the figures."

Often an imperfect performance reveals the intricacy of a dance more clearly than the most polished execution. Not infrequently normal biological processes are first elucidated by a study of the abnormal. Some knowledge of Pathology is necessary for a full appreciation of Physiology.

It is thus very obvious that a beginner, wherever his specialist studies may later lead him, must acquire and retain a knowledge of the fundamental integrated activities of the body as a whole. At the outset of his studies he must be provided with a basic text which will present in true perspective the multifarious activities of the living body.

Unfortunately, specialism has proceeded so far that it is doubtful if any one writer can to-day give a complete and balanced picture of even the elements of Physiology. Collaboration is essential. No happier combination could have been found than that of a physiologist, a biochemist and a clinician. Such successful collaboration ensures that balanced presentation of the knowledge which must be mastered by a beginner.

Because of its importance to Medicine, Physiology once received the title of "Institutes of Medicine." To-day, with her lusty bullying brood of offspring, she justifies that proud title more than ever before. This book is designed to lay a sure foundation for the medical undergraduate. The headlong advances in the Art of Medicine and in the Craft of Surgery in the past few decades are largely attributable to advances in Physiology and in the basic sciences. The medical student neglects his study of Physiology and of Biochemistry at the peril of his future patients.

For those who mean to specialize in Physiology, in Biochemistry, in Pharmacy, or in Pharmacology this book can be no more than an introduction. But it is an introduction which is essential if narrowness in thought and frustration in achievement are to be avoided.

R. C. GARRY.

PREFACE TO THE THIRD EDITION

THE first edition of this book contained 918 pages, the second 1002. This, the third, has 1082 and for that we apologize.

The general arrangement of the book remains unchanged. New information has made necessary considerable alterations especially in the chapters on carbohydrate metabolism, fat metabolism and reproduction. Some additional material on body fluids and electrolytes has been added to chapter thirty-five so as to make what we hope is a more satisfactory, but still elementary, presentation of this subject.

We are well aware that we have, in the course of preparing the third edition of this book, taken perhaps undue advantage of many colleagues with specialized knowledge in different fields. Without their advice we should have floundered more often than we have and we assure them that one of the pleasures we have had in the production of this textbook is in maintaining old friendships and in making new ones. We have in the text, especially in the figures, made individual acknowledgments and we would like especially to thank the following for both criticism and constructive help: Mr B. L. Andrew, Dr G. S. Brindley, Professor W. Burns, Dr F. W. Campbell, Dr S. L. Cowan, Mrs R. S. Creed, Dr E. A. Dawes, Dr G. S. Dawes, Dr P. Fourman, Professor A. C. Frazer, Dr W. C. Hutchison, Dr A. H. James, Dr R. D. Keynes, Dr G. Leaf, Dr J. D. B. MacDougall, Dr H. N. Munro, Dr G. W. Pearce, Dr A. R. Sanderson, Dr R. M. S. Smellie and Dr J. B. de V. Weir. Miss Jean Garvie is again responsible for the index.

Finally, as is customary and right, we record our gratitude to our publishers: to Mr Macmillan whose knowledge of the psychology of authors must be unrivalled; to Mr Parker whose printing programme we repeatedly wrecked; and finally to Mr Matthews whose careful competence adorns many pages.

G. H. B.
J. N. D.
H. S.

October 1956.

PREFACE TO THE FIRST EDITION

THE desirability of collaboration in a book covering a wide field is quite obvious in these days of growing specialization when the many administrative responsibilities increasingly thrust upon teacher and research worker divert them from their proper duties. While no doubt it may be an advantage that the authors should bring experience from very different fields into the writing of one book, yet for any one of a team to satisfy the other members completely is well-nigh impossible and we are now convinced that the difficulties of writing a book increase as the power of the number of its authors.

Now that we see our ideas expressed in cold print we still have sufficient self-criticism left to wonder whether we have in this textbook fully realized our aims.

What then are the aims of this book? The first is that it is intended to be an *introduction* to the study of physiology and biochemistry. Although it is intended primarily for medical students it should also serve the needs of those studying dentistry, veterinary science, or pharmacy, as well as students of pure science in the early stages of their study of physiology and biochemistry. We hope that we have included sufficient material of an applied or clinical nature to emphasize the importance of physiological knowledge in clinical medicine and we trust that the medical student will find it profitable to retain the book and consult it from time to time in his clinical years.

The custom of teaching the elementary stages of physiology and biochemistry in close association has more than mere convenience to commend it, since the subject matter is so inextricably interwoven. This intimate relationship is apt to be forgotten, however, by the more advanced student if he comes to rely on one textbook dealing with biochemistry alone and another confined mainly to physiology. Our aim has therefore been to treat the two subjects as one and to indicate wherever possible their relevance to clinical problems. The numerous cross-references in the text serve to demonstrate the many interrelations at all points of our subject and to underline the importance of considering man as a co-ordinated and integrated whole.

When he has passed the elementary stages and has gained some knowledge of the groundwork of the subject the student can then proceed to more advanced texts and monographs which deal with specialized aspects of physiology or biochemistry. To enable him to do so references are given at the end of each chapter. An additional way of entry into the literature is provided by the references in the legends to those figures which have been taken from original papers.

Because we have the medical student chiefly in mind we have tried to emphasize the human aspects of our subject but it is, of course, quite impossible to confine our attention exclusively to the physiology and biochemistry of man. Indeed any such attempt would leave the student without a sense of obligation to the pioneers of physiology and without any idea of the contribution of animal investigation to the progress of diagnosis and the treatment of human suffering. It is important to remember, however, that clinical investigators have added to physiological knowledge by developing new techniques applicable to man alone. Considerable advances in our knowledge of renal function and of the physiology of the heart and peripheral circulation have arisen in this direct way.

For the reader's convenience in finding figures, tables and plates we have adopted the slightly unconventional method of numbering these in sequence in the chapter to which they belong. Thus Fig. 50, 6, is the sixth illustration in Chapter 50. This system which is now coming into use has we believe much to recommend it. We have adopted the current anatomical terminology and the conventions of spelling and usage advised by the Physiological and Biochemical Societies.

One of our most difficult tasks has been to decide, not so much what should be included, as how much should be left out. It is not possible, nor is it wise, to include the newest work in a book such as this unless its foundation is firm or its relation to other parts of our knowledge is clear; but in trying to effect a judicious selection we realize that we have laid ourselves open to the accusation of failing to be up to date.

As this book has gradually grown we have found ourselves increasingly indebted to many people for advice and help. Since we have tried to acknowledge the source of every figure and table in the text we do not propose to list our benefactors individually here but we should like to tell them again how much we appreciate their kindness and that of their publishers. Some of our friends have done us the service of reading over parts of our manuscript and thus we have been saved from many indiscretions. Without implicating them in any responsibility for the shortcomings of this book we should like to thank particularly a number of colleagues from whom we have had patient criticism and much helpful advice. Professor R. C. Garry has not only read through the manuscript from beginning to end but also readily agreed to write a foreword. Dr H. N. Munro has read nearly all the proofs and has unstintingly given us of his time and knowledge. We are also indebted to Mr B. L. Andrew, Dr G. H. Armitage, Mr F. R. Brown, Dr Ann R. Sanderson and Dr J. B. de V. Weir. Miss M. P. Russell has constructed an index which we hope is free from those irritating frustrations which are so common in indices. Of our publishers, Messrs E. & S. Livingstone Limited, we can only say that we are lost in admiration of their patience and tolerance. Mr Matthews, the artist, has turned out clear and purposeful-looking illustrations from our amateurish scribblings; Mr Parker has carried us over innumerable difficulties with unfailing patience; and we have especially to thank Mr Macmillan, whose eloquence first persuaded us to the indiscretion of writing a textbook and who re-energized us at times when our productivity was declining.

G. H. B.
J. N. D.
H. S.

September 1950.

CONTENTS

PAGE
Foreword v
Preface vii
CHAPTER
1 General Introduction 1
2 Carbohydrates 8
3 Lipids 19
4 Some Physico-Chemical Considerations 29
5 The Proteins 40
6 Nucleotides, Nucleic Acids and Nucleoproteins . . 57
7 Enzymes 69
8 Biological Oxidations and Reductions 80
9 Water and Minerals 93
10 The Vitamins 109
11 Energy Exchange 152
12 Food and Nutrition 173
13 Temperature Regulation 198
14 Hunger and Thirst 214
15 Mouth, Œsophagus and Deglutition 218
16 The Stomach 230
17 Digestion and Absorption in the Small Intestine . . 251
18 The Large Intestine 276
19 Intermediary Metabolism Methods of Study . . . 283
20 Carbohydrate Metabolism 288
21 Fat Metabolism 323
22 Protein Metabolism 335
23 The Blood 360
24 The Formed Elements of the Blood 385
25 Blood Pigments 404
26 Summary of Liver Functions 418
27 Circulation 422
28 The Heart 424
29 Dynamics of the Peripheral Circulation 480
30 Vasomotor Control 507
31 Circulation through Lungs, Liver, Spleen and Brain . 527
32 Respiration 536
33 Respiratory Function of the Blood 571
34 The Kidney 590

CHAPTER		PAGE
35	THE FLUIDS OF THE BODY	629
36	SPECIAL SENSES	648
37	THE SKIN	651
38	THE CHEMICAL SENSES	661
39	VISION	669
40	SPEECH AND HEARING	705
41	NEURONE AND SYNAPSE	733
42	MUSCLE	771
43	AUTONOMIC NERVOUS SYSTEM	804
44	NERVOUS SYSTEM	817
45	SPINAL CORD	822
46	THE BRAIN STEM AND POSTURAL REFLEXES	847
47	THE CEREBELLUM	865
48	THE DIENCEPHALON	876
49	THE TELENCEPHALON	883
50	CONDITIONED REFLEXES	911
51	THE ENDOCRINE GLANDS	918
52	REPRODUCTION	948
53	THE PITUITARY BODY	998
54	CELL DIVISION AND HEREDITY	1013
55	GROWTH AND SENILITY	1021
	UNITS AND MEASURES WITH CONVERSION FACTORS	1026
	STANDARD WEIGHTS	1028
	INDEX	1029

CHAPTER 1

GENERAL INTRODUCTION

IN the subjects of Physiology and Biochemistry we are concerned with living matter at two different levels. The biochemist studies biological systems at the molecular and atomic level, while the physiologist is concerned with the intact organ, or the whole organism. Both have a dynamic outlook, and indeed biologists in general at the present time are more interested in the function of living matter than in its structure. Physiologists and biochemists are concerned essentially with changes in the organism as it reacts to changes in its environment. In other words, they try to take a cinematographic view of life, although it may be necessary to stop the cine-film from time to time to make a detailed examination of a single frame. For example, the chemical analysis of a tissue is in fact a statement referring to one particular instant, but a series of such analyses indicates a progressive change and thus allows the biochemist to build up a dynamic picture of the chemical activities of the body. Similarly the information obtained by the pathologist from a microscopic examination of a stained section of a tissue may be of value in indicating how the minute anatomy of the part differs from normal. From such an examination, however, it is often difficult or impossible to draw any firm conclusion as to the function of the tissue. Such a section indeed represents only a momentary glimpse of what may be a continuously changing process. One must recognize too that disturbances of function occur that seem to have no structural basis. For example, interference with an enzyme system may itself produce disease, and changes in the arrangement of molecular layers may alter the properties of cells and tissues.

In this book the terms 'physiology' and 'biochemistry' are used in a relatively narrow sense. We shall deal, so far as we can, with human physiology, supplementing it by descriptions of mammalian experiments and by occasional but important references to the frog. Anyone actively engaged in the study of physiological problems is soon made painfully aware of minor differences of function between different species and is always reluctant to assume that what holds good for one animal applies without modification to man. There are, however, obvious limitations to experiments on man and it will always be necessary in a text-book such as this to fill in some gaps by reference to lower animals.

The single cell is the functional unit of the body and each tissue is made up of vast numbers of cells. We must not imagine, however, that a single cell is simple; it is only simple in the sense that there is little visible difference between one part of its cytoplasm and any other and we cannot localize the places at which any particular chemical change occurs. Thus, even if we take the smallest living functional unit, it is difficult to describe the characteristics and behaviour of living material and to say how it differs from non-living material.

In the case of the more complex animals such as birds and mammals, the first characteristics of life that come to mind are warmth and movement. If an animal is still and cold it is assumed to be dead. By taking in food and oxidizing it an animal obtains energy which is used to produce heat or movement. The energy is obtained by oxidation of body constituents by a process known as *catabolism*. The energy obtained by these catabolic processes may also be used for *anabolic* or synthetic processes such as are necessary for growth. Since both processes occur side by side it is convenient to use the word *metabolism* when referring to the total chemical changes occurring in the cell or in the body. Living matter is not a substance of definite chemical constitution but, as Noël Paton put it, "It might be compared to a whirlpool constantly dragging things into its vortex and constantly throwing them out more or less changed but itself continuing apparently unchanged throughout." It has recently become possible by labelling molecules with isotopes to trace their movements and to confirm experimentally this generalization of the older physiologists; actually the whirlpool moves even faster than they might have been tempted to speculate. So long as metabolic processes continue, however slowly, the cell is alive; their arrest is death. Since these chemical processes are under the control of enzymes the cells can move or grow only within certain limits of temperature. If the temperature is too high the enzymes are destroyed, while at low temperatures enzymatic reactions are retarded and finally cease.

Living material is *organized*, that is, it has a definite structure. Moreover, particular functions, such as movement or secretion or conduction, are carried out by parts whose structure is peculiarly fitted to these purposes. At once, however, we have to make reservations since, if we leave the higher animals and consider a single cell, we find as already mentioned little differentiation in its protoplasm in spite of the multiplicity of its functions. Organization may not, therefore, be a fundamental characteristic of living material. It is difficult to say whether certain plant viruses which can be obtained as crystalline nucleoproteins are alive or not. Although they possess the power of growth and reproduction within a living cell the fixed and characteristic arrangement of their atoms in space is reminiscent rather of the non-living than of the living world.

Growth of a single cell cannot go on indefinitely because, as the cell increases in volume, its surface through which oxygen and food materials are admitted becomes too far removed from its centre, so that the supply of essential material to the latter is endangered. Before this stage is reached the cell divides into two daughter cells, a process known as *reproduction*.

Because a cell does not live in isolation, and is dependent on obtaining its food from outside, it must be capable of reacting to changes in its environment. Such changes are called *stimuli*. If a stimulus increases the rate of chemical changes in the cell it is said to *excite*; if it decreases the metabolic rate it is said to depress, or to be an *inhibitory* stimulus.

For a cell to survive there must be integration of function within it. In multi-cellular organisms there must be *co-ordination* of the activities of various cells, either by chemical messengers (*hormones*) or by a system of *nerves*.

Living organisms possess two properties which at first sight seem to conflict. These are best described under the headings of *adaptation* and *homoiostasis*. Simple forms of life can survive over a wide range of temperature and can adapt themselves to changes in their environment and to the foodstuffs available. Indeed, if they could not do so they would soon die. The study of adaptation forms a large part of the subject of physiology, for the cells of the body can adjust themselves to a wide variety of changes. On the other hand, many physiological reactions are directed towards preserving the *status quo*. Practically all the cells in the body except those on the surface are provided with a fluid environment of relatively constant temperature, hydrogen ion concentration and osmotic pressure. This permits many bodily activities, the functioning of the nerve cells and the working of enzymes for example, to be carried out under optimum conditions. As Claude Bernard put it: "La fixité du milieu intérieur est la condition de la vie libre, indépendante." Very small changes in the *milieu intérieur*, often too small to be measured with physical apparatus, produce alterations of breathing, of renal function, or of the circulation, which quickly restore the internal environment to its original state. Many of the adjustments of the activity of the cells and organs of the body are examples of the working of this principle, which Cannon called homoiostasis. The accuracy of regulation of the internal environment is greater in man than in the lower animals and Barcroft believed that the acquisition of a more stable environment for the brain had contributed greatly to his mental superiority.

Although the regulation of the internal environment reaches its highest development in man he has gone still further by attempting to control his external environment. At first this involved the wearing of clothing, the building of shelters and houses, and the use of artificial heating, lighting and ventilation. These have not all proved to be unmixed blessings. Artificial lighting, for example, while it makes life more pleasant in the darkness of winter, introduces new problems of eyestrain, fatigue and extended hours of work. Later, with the advent of the industrial era, unfavourable environments were created in which men have to labour, but the dangers of many of these were mitigated by the design of machinery which could be easily and safely worked. More recently man has attempted to conquer the air, an adventure which has resulted in a vast number of new biological problems. The description and investigation of such problems falls properly under the heading of applied physiology, but there is no definite boundary between pure and applied in this or any other science.

Although our knowledge of the properties and functions of living cells is incomplete, yet we know that they cannot transcend the laws of physics and chemistry. It is thoroughly well proved that the laws of conservation of matter and energy apply to the animal body. Investigation of living matter is largely a question of observation supplemented by the methods of physics and chemistry. Thus we may measure pressures and potentials, make chemical analyses, or trace the pathways of chemical substances through the body. The results of these observations are correlated, interpreted in the light of previous knowledge, and used as evidence for or against a particular hypothesis. Thus physiology and biochemistry imply much more than a mere catalogue of the functions of the various parts of the body. Since neither has yet become a highly organized body of knowledge comparable to the exact sciences of physics or mathematics both contain much which cannot yet be given its full interpretation or significance. Although such information may appear irrelevant, it must always be borne in mind that empirical findings may have immediate practical importance in the diagnosis or treatment of disease.

At one time the organic basis for a patient's symptoms could be established only at post-mortem examination. The trend in modern medicine and surgery, however, is to study the living patient more and more intensely in order to understand not only his symptoms but the way in which normal physiological and biochemical processes have broken down, for disease is coming increasingly to be thought of as a disordered physiology in which the compensatory mechanisms have been overstrained or are inadequate. The cure and prevention of many diseases is due to the rapid advances in our knowledge of bacteriology and immunology but there are many morbid processes not in this category and we find that the methods of experimental physiology are being increasingly applied to the study of clinical problems. Biochemical methods have been used for many years for the examination of the body fluids, to provide important diagnostic information and for the control of treatment. Thus, for example, much of our recent knowledge of cardiac diseases, cardiac failure and the surgical treatment of congenital cardiac malformations is based upon physiological studies made in the laboratory or in the clinic, while the modern treatment of diabetes mellitus depends almost entirely on material and methods originally developed in the laboratory.

If research in medicine and surgery is to prosper its exponents must have considerable knowledge of the basic sciences and constant co-operation between the clinician and the laboratory worker is essential. The solution of many clinical problems will no doubt come from the laboratory, because many of the physiologist's or biochemist's attacks on more fundamental questions bring results which are applicable to man. At the same time the study of disease has at times added to and clarified studies made in experimental laboratories. The increasing importance of a close connexion between work in laboratories and clinics emphasizes the value to a doctor of a sound training in physiology and biochemistry.

The bodily activities are so closely dependent on one another that the workings of a part of the body cannot be comprehended without an understanding of the functioning of the whole. For example, in describing the activities of the heart we have to discuss the influence of the peripheral blood vessels, of the central nervous system, of respiration and of the chemical changes occurring in cardiac muscle. Our subject may, therefore, be compared to a circle in which all the parts are connected together, no one portion being more important than another. It is difficult, therefore, to know where to enter the circle to begin our study, for it is only when we have completed the circle that we can understand our subject in its relation to the whole. For this reason it is necessary to consider briefly the subject as a whole before beginning a more detailed description of its various parts.

The source of all the energy required by the body for carrying out muscular activity, for respiration, for the beating of the heart, is the food. This consists mainly of proteins, fats and carbohydrates which are oxidized (burnt) in the tissues. In addition to sources of energy the food must contain inorganic substances which are necessary to make good the loss of salts in the excreta and to provide material for formation of blood and bone. The food must also supply certain substances which the body cannot synthesize, such as vitamins and essential amino-acids. Since there is continuous loss of fluid, by way of the kidneys as well as by the skin and lungs, water must be drunk to make good this loss. When food is swallowed it reaches the stomach and small intestine, where it is broken down by enzymes into substances of relatively simple chemical constitution which are absorbed through the lining of the small intestine into the blood stream and distributed throughout the body. In the case of fats, however, absorption may occur with very little modification of the original chemical structure.

The oxygen required for combustion of the foodstuffs reaches the blood through the lungs. During breathing the chest expands and air rushes into the lungs which are spongy organs richly supplied with blood vessels. Oxygen diffuses readily through the very thin walls of the capillary vessels of the lung tissue, becoming attached to the hæmoglobin contained in the red cells in which it is distributed throughout the body by the circulation. The carbon dioxide produced in combustion in the tissues is taken up by the blood and carried to the lungs where it escapes from the blood and is breathed out. By-products of oxidation not needed by the body reach the kidneys in the blood and are excreted into the urine.

The heart is a two-sided muscular pump which drives the blood along the blood vessels. The left side pumps blood to the heart muscle itself, to the brain, skeletal muscles, kidneys and intestine. The blood from these parts returns to the right heart which sends the blood to the lungs where oxygen is taken up and carbon dioxide is eliminated. The oxygenated blood then returns to the left side of the heart and is pumped out to the tissues. The blood is conveyed away from the heart at a fairly high pressure in thick-walled tubes, the arteries. These vessels branch repeatedly and become smaller in diameter, with thinner and thinner walls. In the tissues the smallest blood vessels, the capillaries, are bounded by a single layer of cells through which gases, fluid, or chemical substances of small molecular size move easily. The blood is drained away from the tissues at low pressure in wide vessels (the veins) with relatively thin walls.

The skeletal muscles are the main effector tissues. By their contractions the position of the bones is altered and respiration and speech are made possible. The highly complex movements of the limbs in walking, and of the tongue in speech, are co-ordinated by the central nervous system, consisting of the brain and spinal cord. Nerves called *efferent* or *motor* nerves leave this system and pass to all the structures of the body and control muscular movement as well as the secretion of the glands, the heart beat and the calibre of the blood vessels. Central control is, however, of no value unless the centre has full information about events in the body and around it. This information is conveyed to the central nervous system by the *sensory* or *afferent* nerves which carry impulses from the eye, the ear, the skin, the muscles and joints, and the heart and intestines. The sensory nerves are actually much more numerous than the motor nerves. Although many of the activities presided over by the central nervous system are exceedingly complex, relatively few rise to consciousness. We are quite unaware, for example, of the muscular adjustments needed to maintain balance or to move our eyes so that images of the external world are kept fixed on the retinæ. These adjustments are called reflex and the pathways involved, namely sensory nerves, central nervous system and motor nerves, are called *reflex arcs.* The lowest part of the brain, the medulla oblongata, is responsible for the muscular movements of respiration, for the control of the heart rate and the regulation of the blood vessels, and is concerned in the maintenance of posture. The cerebellum, which lies above the medulla, is concerned with co-ordination of muscular movements. The fore-part of the brain, the cerebrum, has a layer of grey matter (nerve cells) on its surface and also masses of grey matter within. The grey matter is interconnected by innumerable nerve fibres which together make up the white matter. The cerebrum is concerned in all the higher mental activities and with reading, writing and speaking, as well as in so-called voluntary movements, the perception of touch and temperature and in the special senses of vision and hearing.

In addition to the rapidly acting co-ordinating and integrating mechanism of the nervous system there is a chemical (*humoral*) system which responds more slowly. For example, during the digestion of food in the duodenum a chemical substance (*hormone*) called secretin is produced in the mucous membrane, absorbed into the blood and carried to the pancreas, which responds by pouring out its digestive juices. The pancreatic secretion can, however, also be controlled through the nervous system. It is probable that in the case of the pancreas activity is initiated by nerve impulses but is maintained through the chemical agency. The thyroid gland in the neck produces a chemical substance which is absorbed directly from the gland into the blood stream and regulates the rate of metabolic activity of the tissues. Again, the internal secretion from special cells of the pancreas, the hormone insulin, influences greatly the body's ability to utilize carbohydrate. Co-ordination of the activities of all the ductless (endocrine) glands is brought about chiefly through the anterior lobe of the pituitary body. This small organ apparently produces a number of hormones, some of which control endocrine glands (thyroid, suprarenals, gonads), while others influence metabolism more directly.

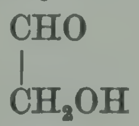
Finally, under the heading of reproduction, we shall have to consider all those processes which are necessary for the maintenance of the species. The male cells, the *spermatozoa*, are produced in the testis, and when deposited in the female genital tract may fertilize an *ovum* produced in the ovary. In mammalia a series of complicated changes, mainly under hormonal control, occurs in the uterus to provide for the growth of the fertilized cell. By repeated division the fertilized ovum develops into the embryo whose nutrition in the uterus is carried out by transfer of materials across the placenta. At the end of pregnancy the muscular wall of the uterus contracts, the fœtus is delivered and then acquires oxygen directly, by breathing air into the lungs, instead of indirectly through the placenta. In the meantime the mother's mammary glands have enlarged in preparation for lactation and soon after parturition they produce milk for the nourishment of the infant. Many of the organs of the infant, especially the central nervous system and the kidneys, are somewhat undeveloped at birth, but as he grows these disabilities disappear and he is then able to lead an active independent existence. An important landmark in development is puberty. At this time the gonads complete their development and in so doing produce physical and mental changes characteristic of the adult. The ovaries continue to produce ova till the menopause in the fifth decade, but the production of spermatozoa by the male is carried on much longer. With the approach of old age the arteries harden, the lens of the eye and probably many other tissues become less elastic, muscular power declines and there is probably some deterioration of mental activity. If the hazards of accident, infection and malignant disease are avoided, death occurs in old age by a gradual process of degeneration.

Having provided in these general terms a conspectus of the contents of the chapters which follow we must now go closer and view each part of the subject in greater detail. It is convenient to begin by dealing with the chemical composition of living matter and of the food we eat.

CARBOHYDRATES

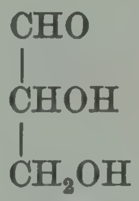
THE carbohydrates are usually defined as substances containing carbon, hydrogen and oxygen, the hydrogen and oxygen being present in the same ratio as in water. Their general formula is, therefore, $C_m(H_2O)_n$. Although satisfactory for most purposes this definition is not strictly accurate, since some carbohydrates are known in which the proportion of oxygen is lower than that required by the general formula.

The easiest approach to the study of their structure follows from the consideration of all carbohydrates as either hydroxy-aldehydes or hydroxy-ketones. The simpler carbohydrates are known as sugars. Those containing an aldehyde group are known as aldoses, and the simplest of these is hydroxy-acetaldehyde. The aldose sugar containing three carbon atoms is glyceric aldehyde or glycerose.



$$\begin{array}{l} CHO \\ | \\ CH_2OH \end{array}$$

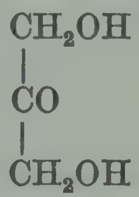
hydroxy-acetaldehyde



$$\begin{array}{l} CHO \\ | \\ CHOH \\ | \\ CH_2OH \end{array}$$

glycerose (glyceric aldehyde)

Ketoses are sugars containing a ketone grouping instead of the aldehyde grouping and the simplest of these is dihydroxy-acetone.



$$\begin{array}{l} CH_2OH \\ | \\ CO \\ | \\ CH_2OH \end{array}$$

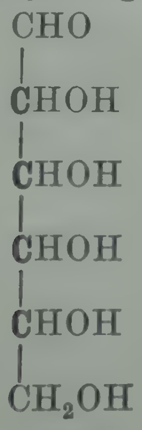
dihydroxy-acetone

Glycerose and dihydroxy-acetone both contain three carbon atoms and are therefore termed trioses. Tetroses contain four carbon atoms, pentoses contain five, and hexoses contain six carbon atoms. Such simple sugars are also known as *monosaccharides*.

When two monosaccharide units condense together with the elimination of water, a *disaccharide* is formed; when three condense together the product is a *trisaccharide*; while the product formed by the condensation of a large number of monosaccharide units is known as a *polysaccharide*.

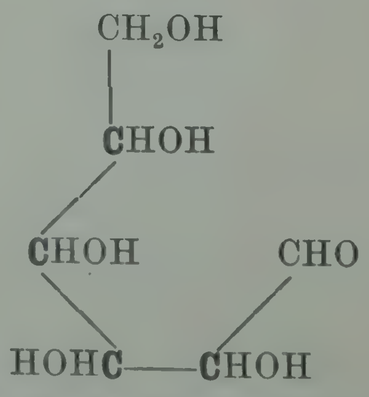
MONOSACCHARIDES

The simplest hexose in the aldose series (aldohexose) might be expected to have the following structure in which the six carbon atoms form a chain, with the aldehyde group at the end.



$$\begin{array}{l} CHO \\ | \\ CHOH \\ | \\ CHOH \\ | \\ CHOH \\ | \\ CHOH \\ | \\ CH_2OH \end{array}$$

or

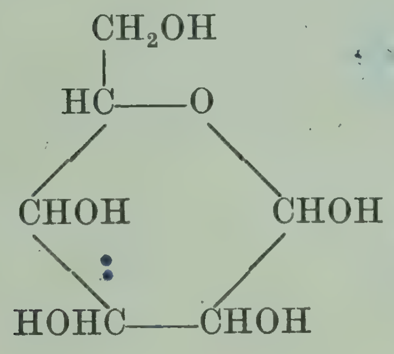


The reason for folding the chain as shown on the right will appear later.

There are four asymmetric carbon atoms in this structure; the aldohexoses are therefore optically active and can exist in several stereoisomeric forms. Thus the aldohexoses include such sugars as glucose, mannose and galactose, each of which can exist in a dextro and a lævo form. Some of the properties of the hexoses, however, are inconsistent with this simple aldehyde structure which, for example, is insufficient to explain the phenomenon known as *mutarotation.*

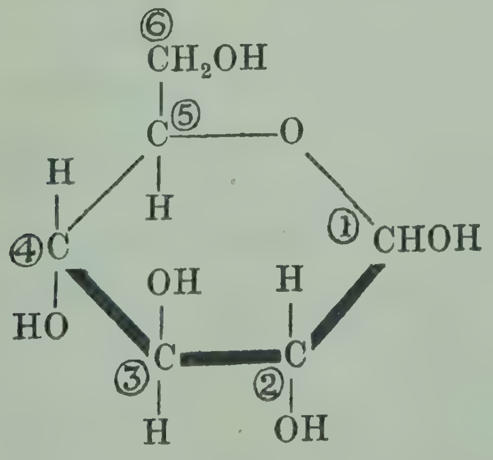
If ordinary glucose is dissolved in water, the solution has a specific rotation of $[\alpha]_D + 111°$, but when the solution is allowed to stand, the rotation falls to $+52{\cdot}5°$. It is also possible to prepare glucose with an initial rotation of $+19°$ which rises however to $+52{\cdot}5°$ when the solution is allowed to stand. It seems, therefore, that glucose can exist in two forms and the change in rotation to a common value, which occurs when either of these two forms is allowed to stand in solution, is known as mutarotation.

To explain this phenomenon and to account for the unduly low reactivity of the aldehyde group in glucose, it is accepted that the carbon atom of the aldehyde group is linked by means of an oxygen atom to the fifth carbon atom in the chain, so that hexoses have a ring structure thus:

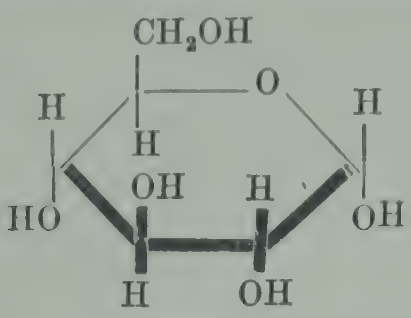


This representation of the hexose molecule in one plane is artificial. If a model is made showing the relationship of the various atoms in space it is seen that five of the carbon atoms and the oxygen atom of the ring form a hexagon, with the other atoms and groups arranged either above or below the plane of the hexagon.

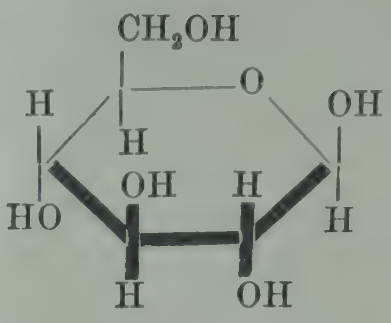
Glucose, for example, can be written in the following way:



It is to be noted that carbon atom number one, which is the carbon atom associated with the aldehyde group, is now asymmetric. Glucose can therefore exist in two forms: one with the hydroxyl group on this carbon atom below the plane of the hexagon and the other with the hydroxyl group above the plane of the hexagon. These two forms known as α- and β-glucose have the following structures:



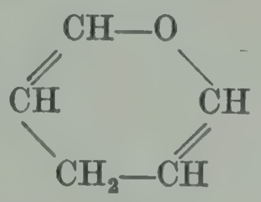
α-glucose



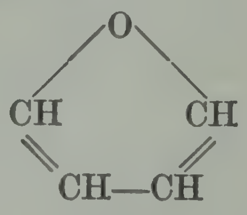
β-glucose

α-Glucose has a rotation of +111° and β-glucose a rotation of +19°. When a solution of either form is allowed to stand an equilibrium mixture of the two forms with a rotation of +52·5° eventually results. Because the equilibrium mixture also contains traces of the open-chain aldehyde form (see p. 8), a solution of glucose gives certain of the reactions characteristic of aldehydes.

When the formula of glucose is written as a six-membered ring it is obvious that glucose is a derivative of the substance pyran; hence sugars containing a six-membered ring system are sometimes known as pyranose sugars. Sugars can also exist in the form of a five-membered ring and may therefore be formulated as derivatives of the parent substance furan. Such sugars are said to have the furanose structure:



pyran

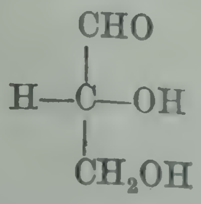


furan

Fructose, for example, exists in the free state as a pyranose sugar, but in combination with glucose as cane sugar (sucrose) it exists in the furanose state (fructofuranose) (p. 14).

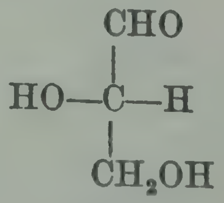
Glucose.—Glucose (dextrose, grape sugar) is a white, crystalline solid, easily soluble in water. Like all sugars it has a sweet taste. It occurs in the animal body and is present in blood in a concentration of 0·1 per cent. Glucose also occurs in nature in the combined form as polysaccharides which on hydrolysis yield free glucose. For example glycogen, or animal starch, which is found in the liver and in the muscles, is built up from glucose units and so are the plant polysaccharides starch and cellulose.

Glucose is optically active and is dextro-rotatory. Configurationally it belongs to what is known as the "D" series of sugars, that is, sugars in which the configuration at the carbon atom next to that carrying the primary alcohol group is the same as in D-glycerose:



$$\begin{array}{c} CHO \\ | \\ H—C—OH \\ | \\ CH_2OH \end{array}$$

D-glycerose



$$\begin{array}{c} CHO \\ | \\ HO—C—H \\ | \\ CH_2OH \end{array}$$

L-glycerose

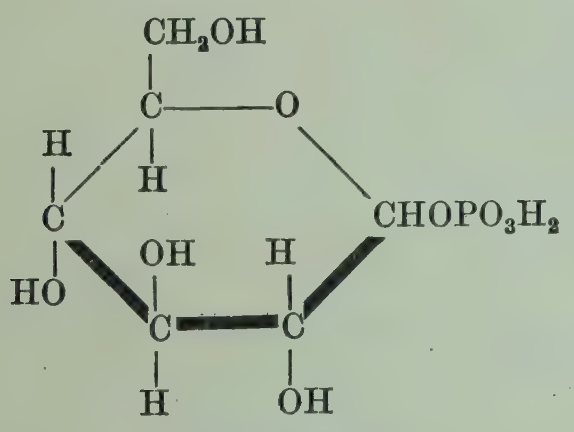
Sugars of the "L" series have the configuration at this carbon atom corresponding to that in L-glycerose. The correct name for glucose is therefore D (+) glucose, the letter "D" indicating the dextro series, and the (+) sign indicating a dextro rotation. (It should be noted that all sugars of the "D" series are not necessarily dextro-rotatory.)

Glucose is a reducing agent, capable of reducing cupric compounds to the cuprous state, and copper-containing reagents such as Benedict's or Fehling's solution are frequently used in testing for the presence of glucose in urine. When either of these solutions is boiled with a solution containing glucose, the cupric copper is reduced to cuprous oxide which appears as a reddish-brown precipitate. Glucose also reduces potassium ferricyanide to ferrocyanide, and this is made use of in the method of Hagedorn and Jensen for estimating the concentration of glucose in blood.

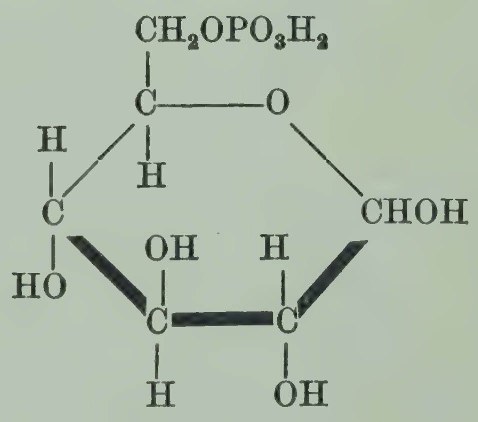
Glucose reacts with phenylhydrazine to form an osazone known as glucosazone. This reaction offers a very convenient means by which glucose can be distinguished from other sugars, since the osazones of the various sugars have different melting-points and show different crystalline forms under the microscope. It should, however, be noted that fructose forms the same osazone as glucose.

Another property of glucose which has been known for a very long time is its fermentation by yeast to yield alcohol and carbon dioxide. This reaction helps to characterize glucose, since some sugars, for example lactose, are not normally fermented by yeast.

Glucose, being an alcohol, forms esters with acids; for example phosphoric acid can react either with carbon atom number one or carbon atom number six in the glucose molecule to yield 1-phosphoglucose, or 6-phosphoglucose. These phosphoric esters of glucose, as we shall see later, are of the utmost importance in carbohydrate metabolism.

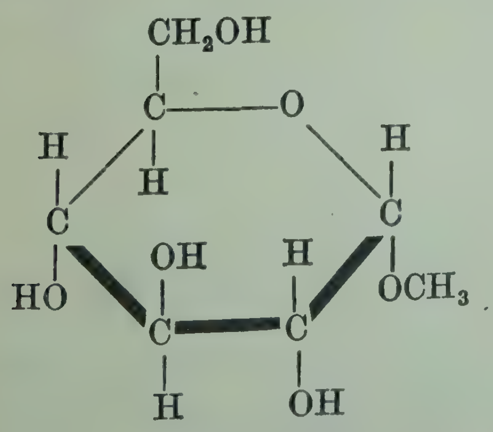


1-phosphoglucose

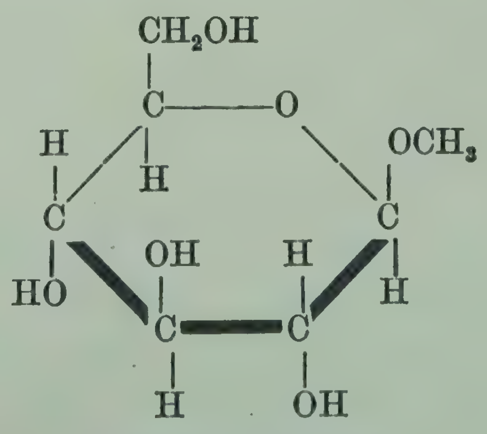


6-phosphoglucose

Carbon atom number one in the glucose molecule is reactive, and the hydrogen atom of its hydroxyl group may be replaced by other radicals, with the formation of compounds known as *glucosides*. For example, it may be replaced by a methyl group with the formation of methylglucosides. There are of course two glucosides, α-methylglucoside and β-methylglucoside, corresponding to the α- and β-forms of glucose.



α-methylglucoside

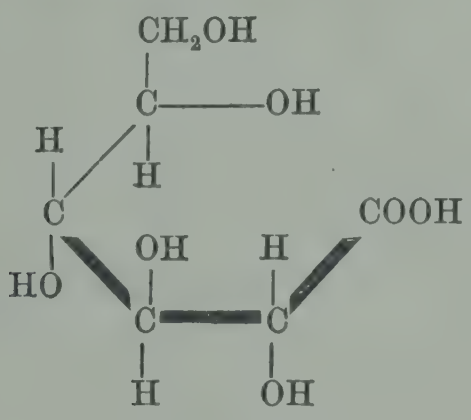


β-methylglucoside

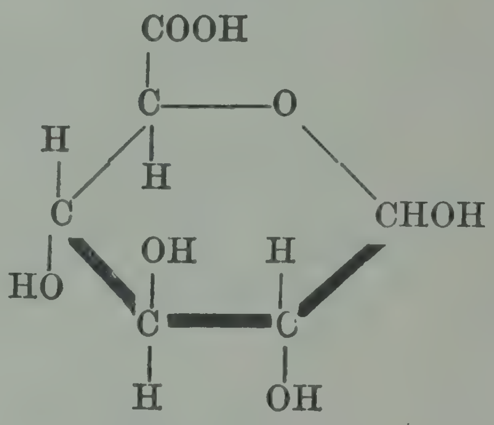
The general term for sugar derivatives of this sort is *glycoside*. Glycosides formed from glucose are *glucosides*, those from galactose are *galactosides*, and so on.

The glycosides are important compounds. Many complex glycosides occur in nature, providing materials of great pharmacological value. For example digoxin, which is extensively used in medicine on account of its action on the heart, is a glycoside.

Carbon atom number one in the glucose molecule may be oxidized to a carboxyl group. When such an oxidation takes place in any hexose the resulting acid is known as *hexonic acid*, and the acid formed from glucose is known as *gluconic acid*. The glucose molecule may, however, also be oxidized at carbon atom six. When a carboxyl group is produced at this carbon atom in any hexose the resulting acid is known as a *uronic acid*, or *hexuronic acid*, and the uronic acid produced from glucose is ***glucuronic acid***.



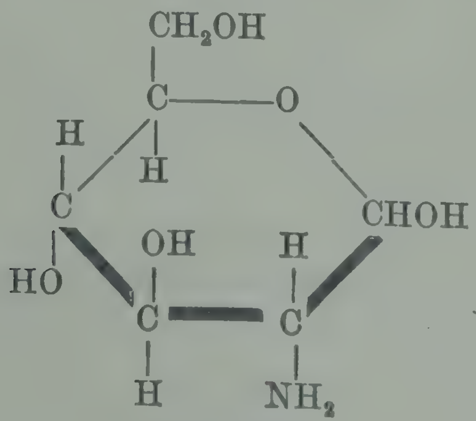
gluconic acid



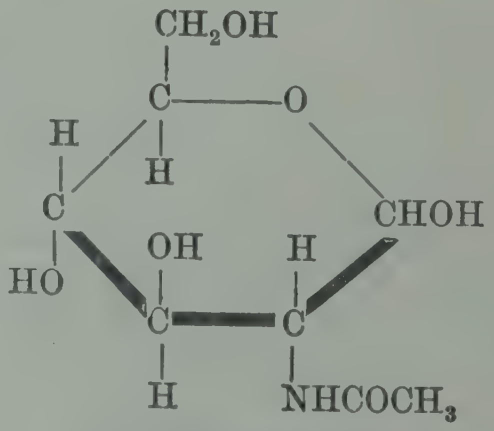
glucuronic acid

The latter is of considerable importance both as a constituent of complex polysaccharides (p. 17), chiefly of bacterial origin, and as a coupling agent. Many substances which are poisonous to the body are excreted in the urine coupled with glucuronic acid, in the form of glucuronides (p. 419).

When an amino group is introduced into hexose sugars the product is known as an *amino-sugar*, or *hexosamine*. The compound formed in this way from glucose is *glucosamine*, or 2-*amino-glucose*, which occurs extensively in nature in complex polysaccharides, usually in the form of its acetyl derivative N-acetylglucosamine.



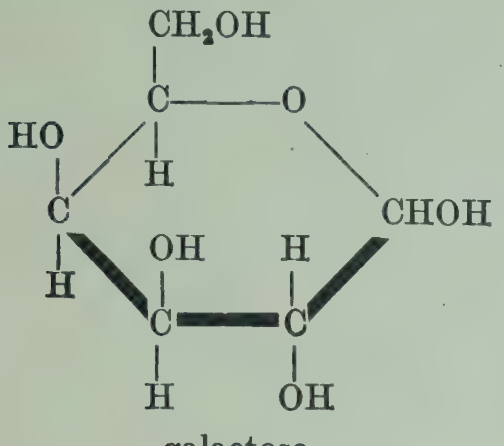
glucosamine



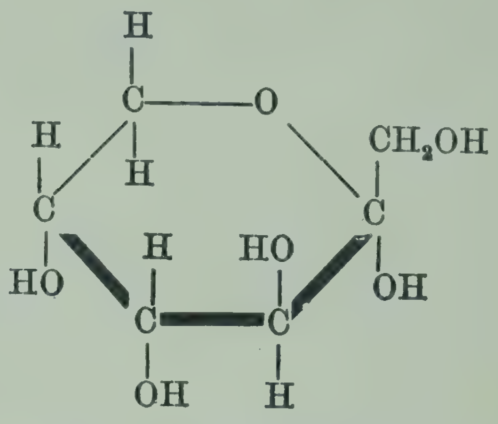
N-acetylglucosamine

Other Monosaccharides

Galactose occurs in nature in the combined form as a constituent of lactose (milk-sugar), in certain complex lipids and in some proteins. It is a stereo-isomer of glucose and is dextro-rotatory. *Mannose* occurs in nature in the form of complex polysaccharides. *Fructose*, sometimes also called *lævulose* or *fruit sugar*, occurs naturally in fruits and honey. In the body it is present in the free state in the seminal plasma and in the blood of the fœtus of ruminants but not of man. Although fructose belongs to the "D" series of sugars it is lævo-rotatory and its correct name is D (–) fructose. It is a ketose sugar and in the free state has a pyranose structure, although when combined in sucrose it exists in the furanose form.

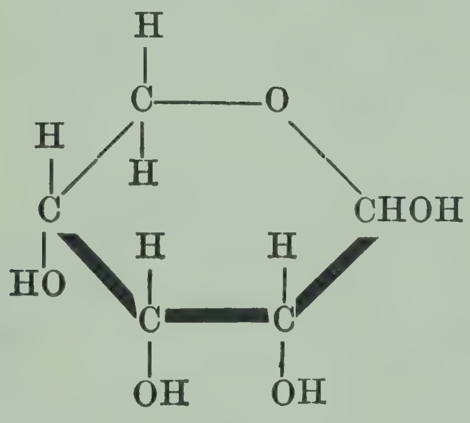


galactose



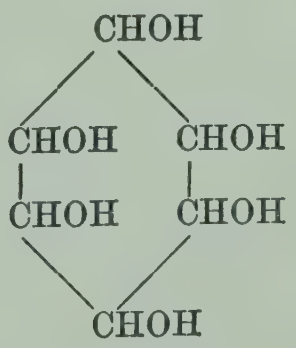
fructose (fructopyranose)

The *pentoses* are monosaccharides with only five carbon atoms in the molecule, and are mostly derived from plant sources. The most important pentose is *ribose* which occurs in the nucleic acids in the tissues (p. 57).



ribose

Inositol (hexahydroxycyclohexane), although not strictly speaking a sugar, is worth mentioning at this stage. It is found in animal tissues such as liver and heart. In plants it occurs chiefly in the form of its hexaphosphate which is called *phytic acid*. It is also found in brain tissue in certain complex phospholipids (p. 23).

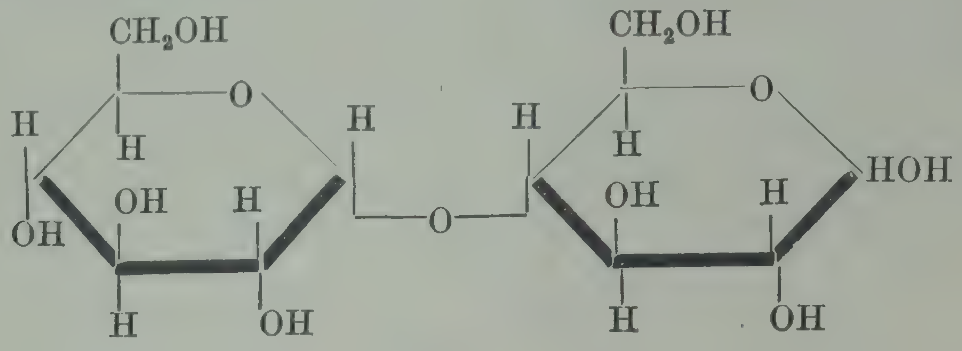


inositol

DISACCHARIDES

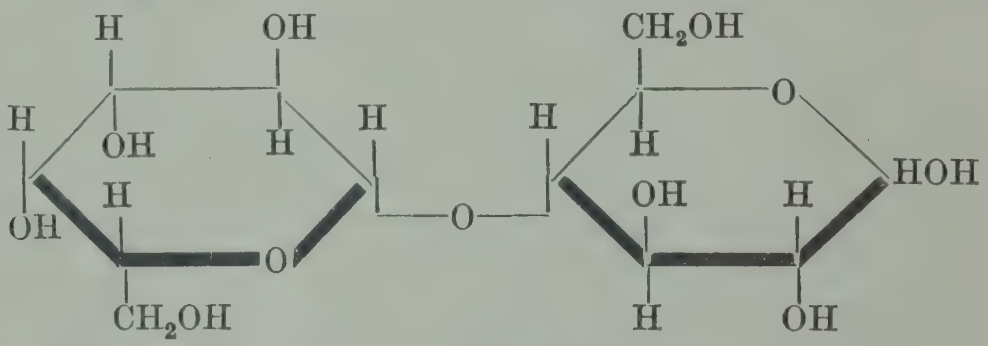
The disaccharides consist of two molecules of monosaccharides condensed together with the elimination of water. There are three important disaccharides, *maltose*, *lactose* and *sucrose*. Maltose is composed of two molecules of glucose, lactose of one molecule of glucose and one of galactose, and sucrose of one molecule of glucose and one of fructose.

Maltose does not occur free in the tissues but it is important as an intermediate stage in the breakdown of starch to glucose. It consists of two molecules of glucose condensed together by an α-linkage. The reducing group of one of the glucose molecules is involved in the linkage but, since the reducing group of the other is free, maltose is a reducing sugar. It forms a characteristic osazone.



maltose
glucose-α-glucoside

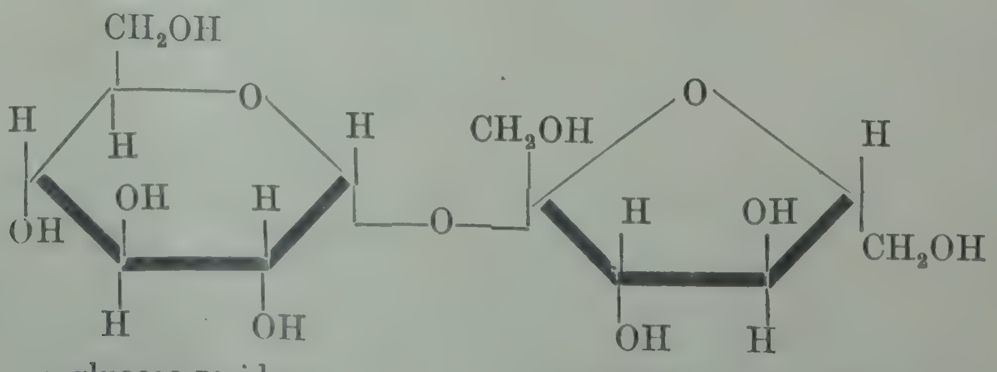
Lactose, or milk sugar, occurs naturally in the milk of mammals and is synthesized in the mammary gland. On hydrolysis it yields one molecule of glucose and one of galactose. It is a β-galactoside and its structure is that of β-galactosido-glucose. Like maltose it has reducing properties.



β-galactose residue
glucose residue

lactose

Sucrose (*cane sugar* or *beet sugar*) occurs naturally in certain plants and is one of the important carbohydrates of the diet. On hydrolysis it yields one molecule of glucose and one of fructose. Since the reducing groups of both the glucose and the fructose are involved in the linkage, sucrose has no reducing properties. When combined in sucrose, fructose exists in the furanose form. Sucrose is dextro-rotatory but the mixture of glucose and fructose produced by hydrolysis is lævo-rotatory, since fructose is more lævo-rotatory than glucose is dextro-rotatory. Hence the rotation is changed from positive to negative during hydrolysis, a phenomenon sometimes referred to as 'inversion.'



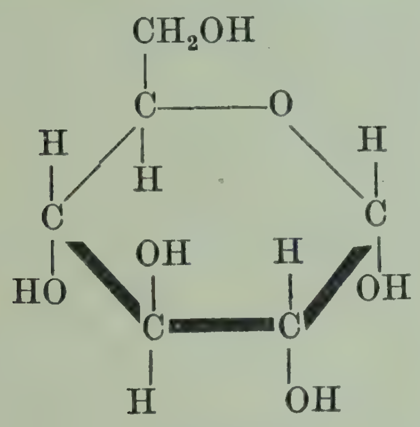
α-glucose residue
fructofuranose residue

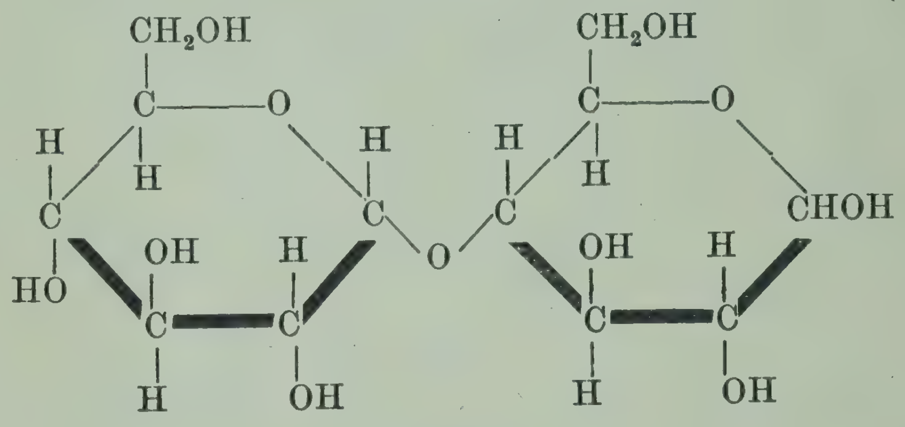
sucrose

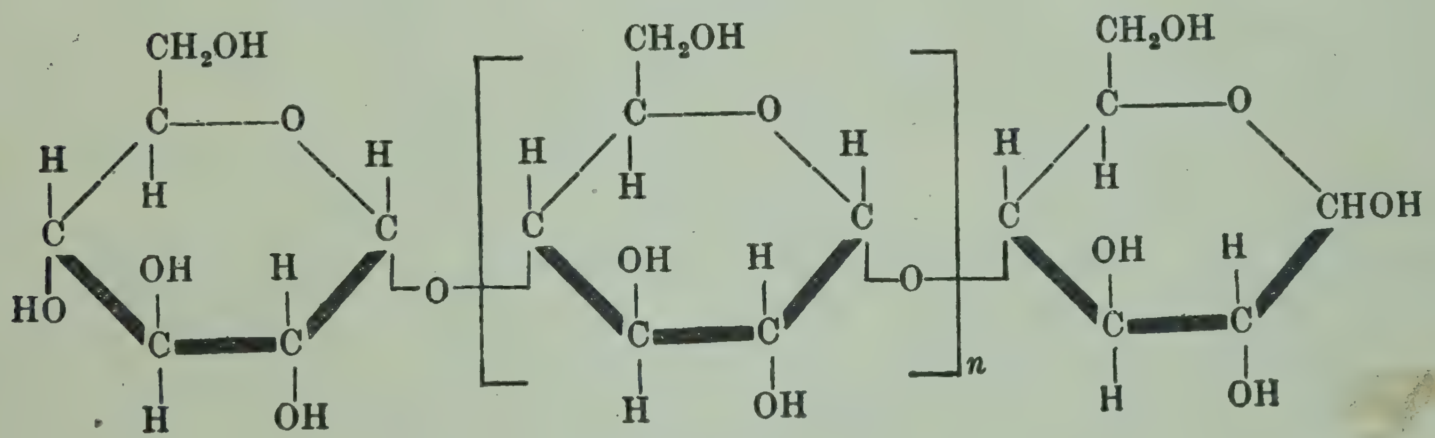
POLYSACCHARIDES

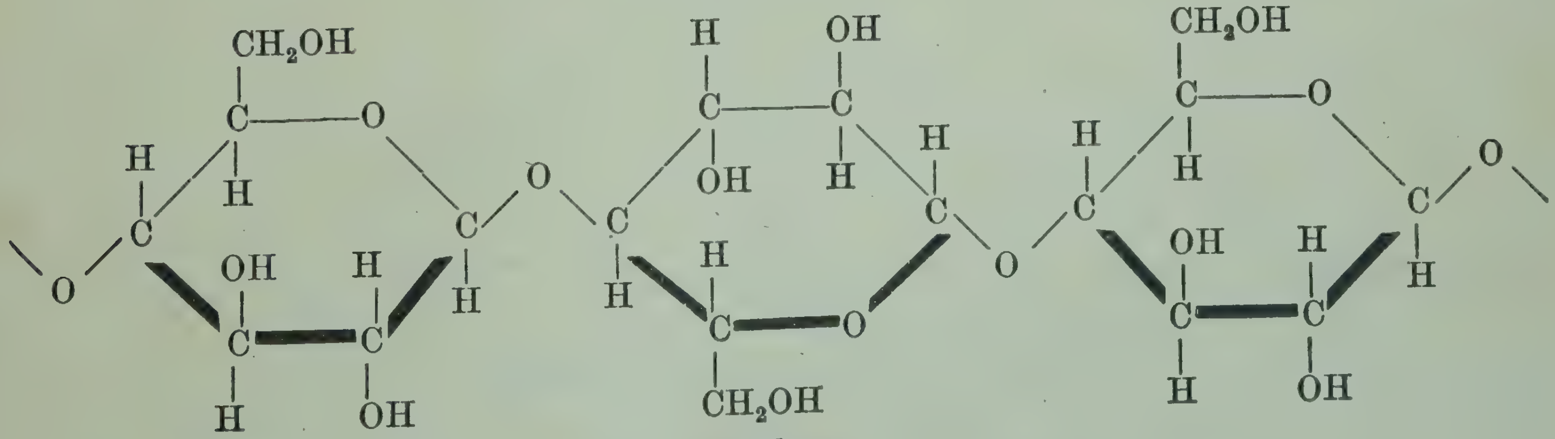
Polysaccharides are formed by the condensation of large numbers of monosaccharide units which are joined together with the elimination of water much in the same way as the amino-acids are joined together to form proteins. Like the proteins, the polysaccharides have high molecular weights and are either insoluble in water or form colloidal solutions.

The three chief polysaccharides, starch, glycogen and cellulose, are built up solely from glucose units. In starch the glucose units are joined together by α-linkages to form chains, which may link up in a laminated or spiral form. Partial hydrolysis of starch yields maltose. In glycogen only 12 or 18 glucose units are present, joined together by α-linkages, and there is evidence that the chain is branched. In cellulose the number of units in the chain may be as many as one or two hundred, and in this case the glucose units are joined together by β-linkages to form the long molecular chains present in cellulose fibres.


α-glucose


maltose


amylose


cellulose

Starch is the form in which carbohydrate is stored in the plant. It occurs in plant tissues, for example the potato, in the form of granules with a thin cellulose coating. These granules themselves are insoluble in water, but on heating the coating ruptures and the starch is liberated. It forms a colloidal solution with an opalescent appearance.

Starch consists of one quarter *amylose* and three quarters *amylopectin*. Amylose consists of long unbranched chains of about 300 α-glucose units forming a helical structure. It gives an intense blue colour with iodine. Amylopectin, on the other hand, is composed of chains of about 22 glucose residues. These chains are arranged in layers one above the other, forming a laminated structure in which the layers are joined here and there by interlayer bridges between carbon atom number 1 in the terminal residue of the upper layer and carbon 6 in one of the residues in the layer below (Fig. 2, 1). Amylopectin gives a red-brown colour with iodine.

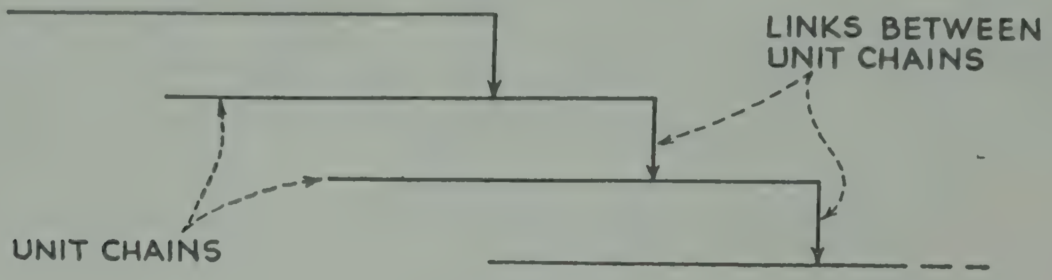


Fig. 2, 1

Plan of a portion of the amylopectin molecule.

Glycogen is the form in which carbohydrate is stored in the animal body. When glucose is to be stored it is converted into glycogen in the liver and muscles and kept there until required, when it may be broken down again into glucose. Glycogen dissolves fairly readily in water to yield an opalescent solution which gives a reddish colour with iodine. The glycogen molecule consists of a large number of chains, each containing 12 or 18 α-glucose units, the chains being linked together to form a branched structure (Fig. 2, 2).

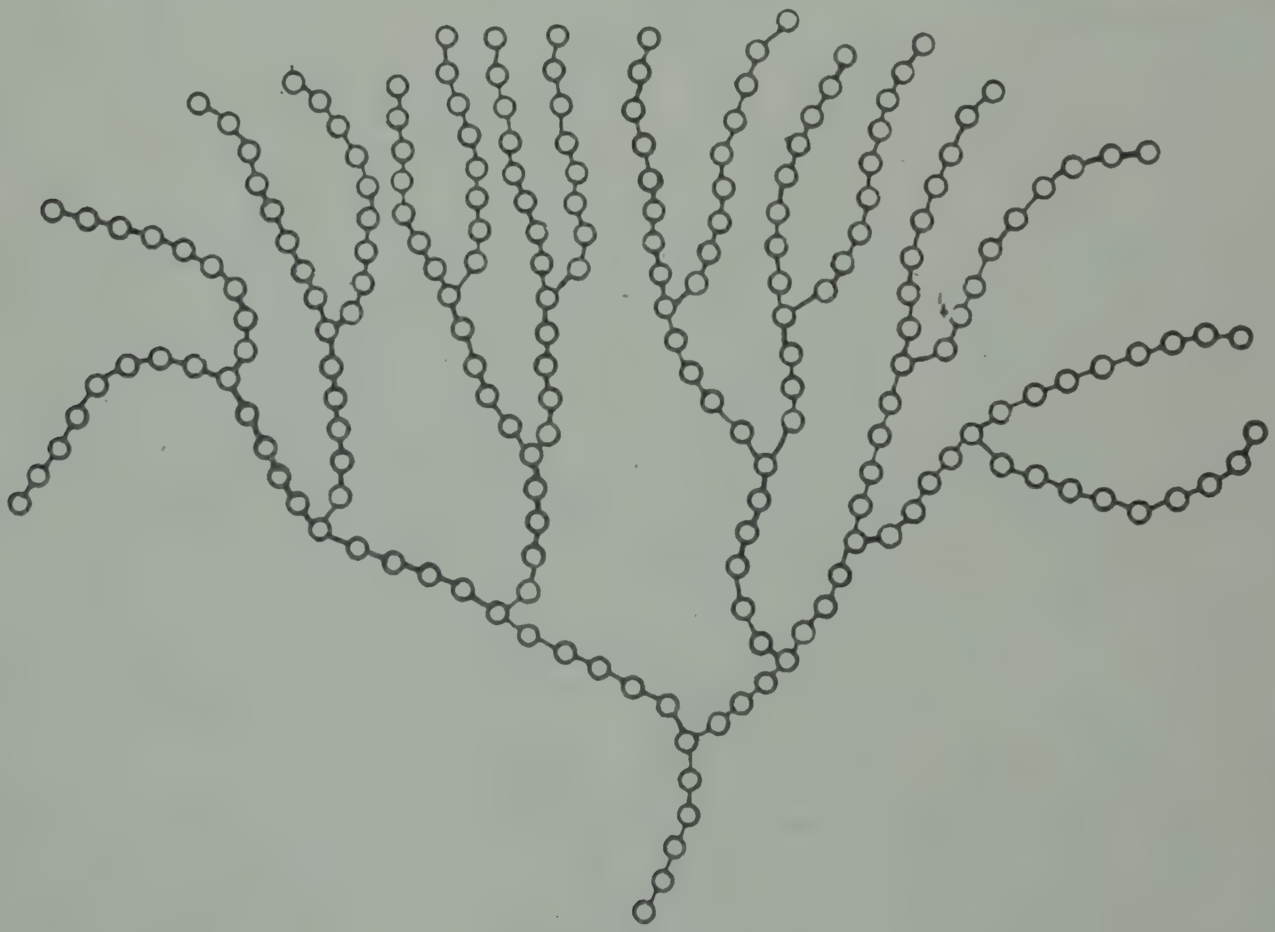


Fig. 2, 2

Plan of a portion of the glycogen molecule.

Cellulose is a very stable insoluble polysaccharide which forms the supporting tissues of the plant. It is insoluble in water and gives no colour with iodine. Although ingested in considerable quantities, it is not digested to any significant extent by man. The herbivorous animals, however, are able to make use of cellulose since the bacteria and protozoa in the rumen or colon convert it, not to glucose, but to small fragments such as short-chain fatty acids, carbon dioxide and methane. The cellulose molecule consists of long chains of glucose units linked together by β-linkages.

Dextrins are intermediate between starch and maltose. They form an ill-defined group of substances which have very feeble reducing properties. The higher dextrins, such as amylodextrin, give a blue colour with iodine. Erythrodextrin gives a reddish-brown colour but achroodextrin gives no colour with iodine. The dextrins form sticky solutions which are used as gums.

Another polysaccharide worth mentioning is *inulin* which is composed entirely of fructose units. It is found in artichokes and in dahlia tubers and is used in a test of renal function.

MUCOPOLYSACCHARIDES

The *mucopolysaccharides* are a group of complex and biologically important materials which exist in nature combined with small, but significant, amounts of protein. The units composing them include, as well as hexoses and pentoses, either uronic acids or hexosamines, or both. The *polyuronides* are composed of uronic acid units and are mainly of plant and bacterial origin. They include many gums, such as gum arabic and gum tragacanth, and also the specific polysaccharides elaborated by certain types of bacteria, such as the *Pneumococci*.

The *polyhexosamines* include the substance chitin which forms the shells of crustaceans and the hard outer portions of insects. It is composed of glucosamine units joined together by β-linkages and is therefore very similar in structure to cellulose. The hexosamines, together with simple hexoses, are constituents of the polysaccharides of the specific blood group substances (p. 381), which are found in the red blood cells and also in some secretions.

Of the polysaccharides containing both uronic acids and hexosamines, one of the most important is *hyaluronic acid* which consists of alternating molecules of N-acetylglucosamine and glucuronic acid. It forms aqueous solutions of high viscosity and is found for example in the skin, in the vitreous humour of the eye, in the umbilical cord and in certain bacteria. It appears to exercise a cementing function in the tissues and probably also in the capillary walls, and it forms a coating gel round the ovum. The synovial fluid, which contains 0·02 to 0·05 per cent. of hyaluronate, owes about 80 per cent. of its viscosity to this substance. Each molecule of hyaluronic acid is broken down by an enzyme called *hyaluronidase* into a number of smaller molecules, with a decrease in viscosity of the solution and an increase in concentration of reducing sugar. This chemical change is accompanied by a striking alteration in the properties of tissues, as the result of which fluid injected into them rapidly spreads from the site of injection. The enzyme is therefore sometimes referred as the 'spreading factor'. It is found in relatively high concentration in the testis and seminal fluid, in the venoms of certain snakes and insects, and in some bacteria.

The sulphuric acid ester of hyaluronic acid is known as *mucoitin sulphate.* It is found in mucous secretions and exercises a lubricating and a protective function. In the intestinal tract, for example, it forms a protective layer over the mucosa which it helps to protect from digestion by pepsin and trypsin. Gastric mucin consists of mucoitin sulphate together with one of the blood group polysaccharides and protein.

A similar material known as *chondroitin sulphate* is found in cartilage, bone matrix and the umbilical cord. It is built up from N-acetylgalactosamine, glucuronic acid and sulphuric acid.

A third mucopolysaccharide containing sulphuric acid is *heparin*, a naturally occurring anticoagulant, which is found in the liver, as its name suggests, and also in lung, spleen, kidney and intestinal mucosa. It can be prepared in the pure state and is used to prevent the clotting of blood (p. 363). Its exact chemical nature is unknown but it contains more than one ester sulphate group.

Mucoproteins or *glycoproteins* are proteins containing small but important amounts of the mucopolysaccharides. Serum albumin, for example, contains small amounts of polysaccharides built up from glucosamine, mannose and galactose, and a similar polysaccharide is found in egg-albumin. Some of the pituitary hormones, such as the follicle-stimulating hormone (F.S.H.) and the luteinizing hormone (L.H.), are glycoproteins. The chorionic gonadotrophins (Chapter 52) are also glycoproteins.

REFERENCES

BELL, D. J. (1952). *Introduction to Carbohydrate Biochemistry.* London: University Tutorial Press.

PERCIVAL, E. G. V. (1950). *Structural Carbohydrate Chemistry.* London: Muller.

CHAPTER 3

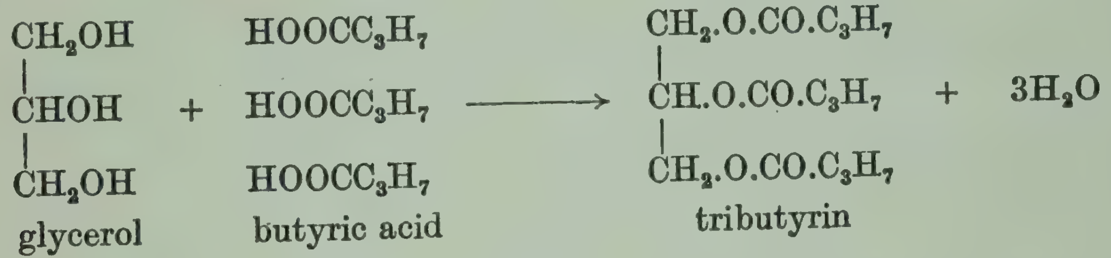
LIPIDS

THE *lipids* form a group of naturally occurring fat-like substances which are all insoluble in water but soluble in the fat solvents such as ether, chloroform and benzene. Lipids are generally esters of fatty acids. They are sometimes referred to as *lipides* or *lipins*, but the term lipin should be used only as a suffix, for example *phospholipin*. The appropriate adjective is *lipoid* or *lipoidal.*

The lipids can be divided into *simple lipids*, including the *fats* and the *waxes*, and the *compound lipids*, including the *phosphatides* and the *cerebrosides.*

THE SIMPLE LIPIDS

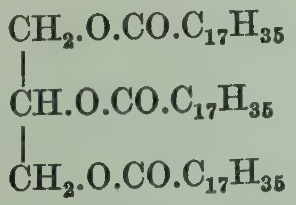
1. **The Fats.**—The fats are all esters of glycerol with fatty acids. Since glycerol is a trihydric alcohol it can react with three molecules of an acid to form a triglyceride. Thus it can react with butyric acid, C_3H_7COOH, to yield glyceryl tributyrate or tributyrin, one of the simplest of the fats, which occurs in butter.



$$\begin{array}{l} CH_2OH \\ | \\ CHOH \\ | \\ CH_2OH \\ \text{glycerol} \end{array} + \begin{array}{l} HOOCC_3H_7 \\ \\ HOOCC_3H_7 \\ \\ HOOCC_3H_7 \\ \text{butyric acid} \end{array} \longrightarrow \begin{array}{l} CH_2.O.CO.C_3H_7 \\ | \\ CH.O.CO.C_3H_7 \\ | \\ CH_2.O.CO.C_3H_7 \\ \text{tributyrin} \end{array} + 3H_2O$$

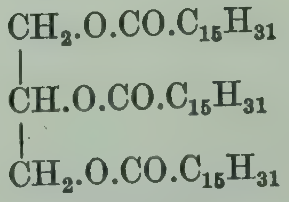
Glycerol reacts with one molecule of a fatty acid to form a monoglyceride and with two molecules to form a diglyceride.

With stearic acid ($C_{17}H_{35}COOH$) glycerol forms glyceryl tristearate or tristearin, and with palmitic acid ($C_{15}H_{31}COOH$) it yields glyceryl tripalmitate or tripalmitin.



$$\begin{array}{l} CH_2.O.CO.C_{17}H_{35} \\ | \\ CH.O.CO.C_{17}H_{35} \\ | \\ CH_2.O.CO.C_{17}H_{35} \end{array}$$

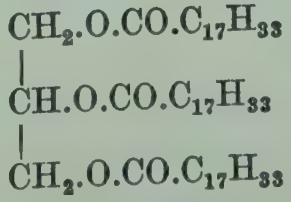
tristearin



$$\begin{array}{l} CH_2.O.CO.C_{15}H_{31} \\ | \\ CH.O.CO.C_{15}H_{31} \\ | \\ CH_2.O.CO.C_{15}H_{31} \end{array}$$

tripalmitin

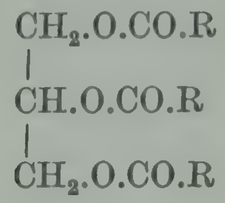
Both tristearin and tripalmitin occur in large quantities in beef and mutton fat. With the unsaturated acid, oleic acid ($C_{17}H_{33}COOH$), glycerol yields glyceryl trioleate or triolein, which is present in olive oil.



$$\begin{array}{l} CH_2.O.CO.C_{17}H_{33} \\ | \\ CH.O.CO.C_{17}H_{33} \\ | \\ CH_2.O.CO.C_{17}H_{33} \end{array}$$

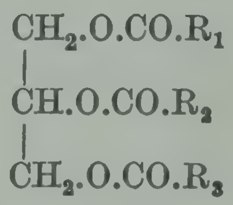
triolein

It will be seen, therefore, that the general formula for a fat is



$$\begin{array}{l} CH_2.O.CO.R \\ | \\ CH.O.CO.R \\ | \\ CH_2.O.CO.R \end{array}$$

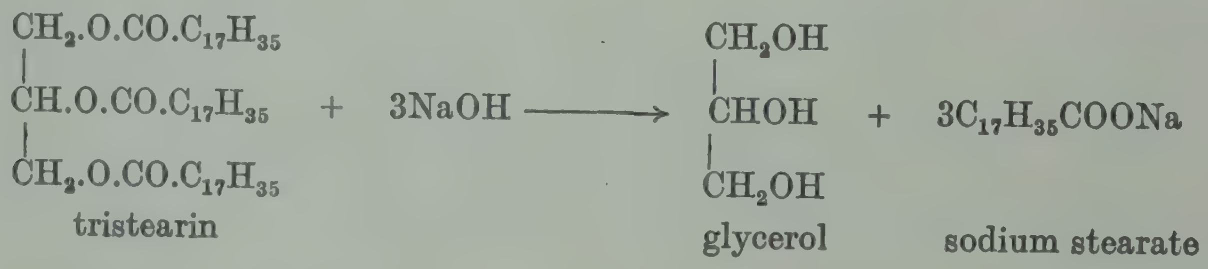
where R varies according to the fatty acid present. It is possible for three different fatty acids to be combined with one molecule of glycerol in a triglyceride, so that the general formula of a fat should more correctly be written thus



$$\begin{array}{l} CH_2.O.CO.R_1 \\ | \\ CH.O.CO.R_2 \\ | \\ CH_2.O.CO.R_3 \end{array}$$

where R_1, R_2 and R_3 may be derived from different fatty acids.

THE PROPERTIES OF THE FATS.—Most fats are greasy solids insoluble in water but readily soluble in fat solvents, including hot, but not cold, ethanol. Some fats, for example triolein, are liquid at room temperature and such fats are popularly called oils, for example olive oil. The term oil is better avoided in this sense, since it is also commonly applied to mineral oils which are not glyceryl esters but hydrocarbons. Some fats are solid at room temperature but are liquid at the temperature of the body; thus, human fat is said to melt about 17°. The melting points of fats are always higher than their solidification points; for example, tristearin melts at 72° but solidifies again at 52°. Fats may be hydrolysed by boiling with an alkali such as sodium hydroxide to yield glycerol and the sodium salts of the fatty acids which are, of course, soaps. This process is termed *saponification.*

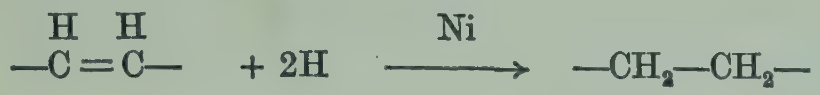


$$\underset{\text{tristearin}}{\begin{array}{l} CH_2.O.CO.C_{17}H_{35} \\ | \\ CH.O.CO.C_{17}H_{35} \\ | \\ CH_2.O.CO.C_{17}H_{35} \end{array}} + 3NaOH \longrightarrow \underset{\text{glycerol}}{\begin{array}{l} CH_2OH \\ | \\ CHOH \\ | \\ CH_2OH \end{array}} + \underset{\text{sodium stearate}}{3C_{17}H_{35}COONa}$$

The *saponification value* of a fat is the number of milligrams of potassium hydroxide required to saponify one gram of fat. Fats may also be hydrolysed by superheated steam, and by enzymes called lipases such as that found in the secretion of the pancreas.

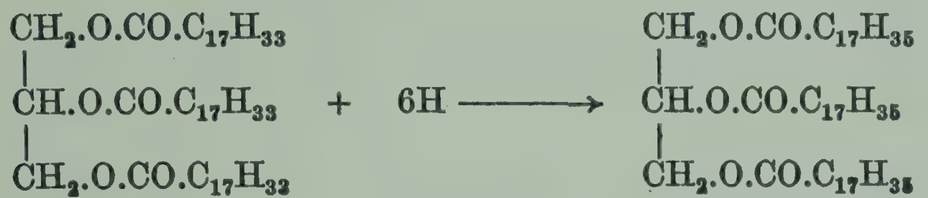
The free higher fatty acids obtained by such hydrolysis have properties which are very similar to those of the fats themselves. While lower fatty acids, like acetic acid, are easily soluble in water, the higher fatty acids, such as stearic acid or palmitic acid, are greasy solids insoluble in water but soluble in the fat solvents. Their sodium and potassium salts, however, are soluble in water and are known as soaps. The corresponding calcium and magnesium salts are insoluble and are precipitated when calcium chloride or magnesium chloride is added to soap solution. Calcium salts of the higher fatty acids tend to precipitate when soap is added to hard water, forming a scum on the surface.

In general, fats containing a large proportion of unsaturated fatty acids have low melting-points and may be liquids at ordinary temperatures. When such fats are treated with hydrogen, in the presence of finely divided nickel as a catalyst, hydrogen is added on at the double bonds and a saturated fat results:



$$\overset{H}{-C}=\overset{H}{C-} \quad + 2H \xrightarrow{Ni} -CH_2-CH_2-$$

In this way tristearin can be obtained from triolein:



$$\begin{array}{l} CH_2.O.CO.C_{17}H_{33} \\ | \\ CH.O.CO.C_{17}H_{33} \\ | \\ CH_2.O.CO.C_{17}H_{33} \end{array} + 6H \longrightarrow \begin{array}{l} CH_2.O.CO.C_{17}H_{35} \\ | \\ CH.O.CO.C_{17}H_{35} \\ | \\ CH_2.O.CO.C_{17}H_{35} \end{array}$$

triolein tristearin

This process is of great commercial and industrial importance since it enables palatable solid fats to be obtained from rather unpleasant greasy liquids. Margarine is a product of this sort. The process of hydrogenation is not usually carried to completion, since it would result in the production of a very hard brittle fat.

Unsaturated fats can also add iodine at the double bonds:

$$-CH=CH- \quad + I_2 \longrightarrow -CHI-CHI-$$

The amount of iodine taken up gives a measure of the degree of unsaturation of the fat. The *iodine value* of a fat is the number of grams of iodine taken up by one hundred grams of fat.

When fats are exposed to the air for any length of time they undergo a certain amount of oxidation, especially at the double bonds, and they also suffer partial hydrolysis with liberation of free fatty acids. A fat thus becomes much less palatable and is usually said to have gone rancid. This process of oxidation can be prevented to a large extent by the presence of small amounts of organic substances which are usually termed *antioxidants*. Many phenols act in this way, and they are of considerable importance in the food industry. Highly unsaturated fats, such as linseed oil, undergo considerable oxidation on exposure to air, forming tough waterproof films. Such oils, called drying oils, are extensively used in the manufacture of paints and varnishes.

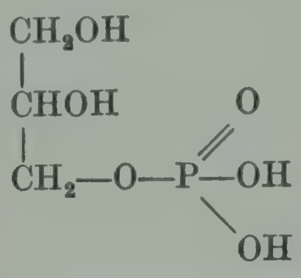
2. **The Waxes.**—The waxes are esters of fatty acids not with glycerol but with complex monohydric alcohols. Beeswax, for example, is an ester of palmitic acid with myricyl alcohol ($C_{30}H_{61}OH$), and spermaceti from the sperm whale is an ester of palmitic acid with cetyl alcohol ($C_{16}H_{33}OH$). Many animal waxes are esters of the steroid alcohol, cholesterol (p. 24).

COMPOUND LIPIDS

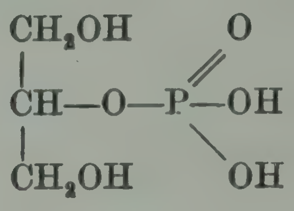
1. **The Phosphatides.**—The phosphatides (phospholipids) are widely distributed in animal tissues and appear to play a very important part in the structure of the cell, particularly the cell membrane, and in the process of fat metabolism. They are especially abundant in brain and

nervous tissues. They frequently exist loosely combined with proteins in the form of *lipoproteins*. Thus a considerable proportion of the protein of the blood plasma is loosely combined with phosphatides. The phosphatides differ from the fats in containing phosphoric acid and an organic nitrogenous base.

(a) *Lecithin.*—The best-known phosphatides are the lecithins which on hydrolysis yield glycerol, fatty acids, phosphoric acid and choline. Partial hydrolysis may yield phosphoglycerol.

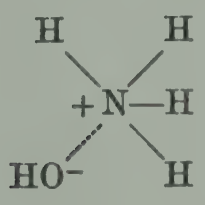

CH₂OH
CHOH
CH₂—O—P—OH
O
OH

α-phosphoglycerol

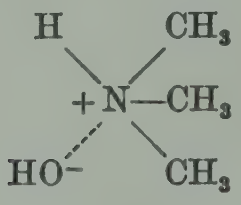

CH₂OH
CH—O—P—OH
CH₂OH
O
OH

β-phosphoglycerol

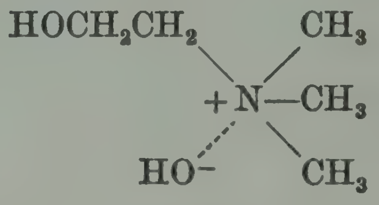
Choline is a derivative of ammonium hydroxide thus:



ammonium hydroxide

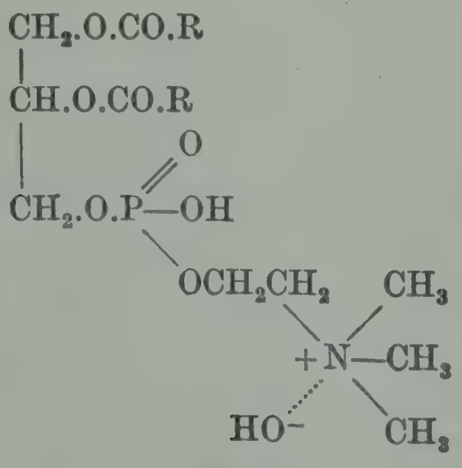

H
CH₃
+N—CH₃
HO⁻
CH₃

trimethyl ammonium hydroxide

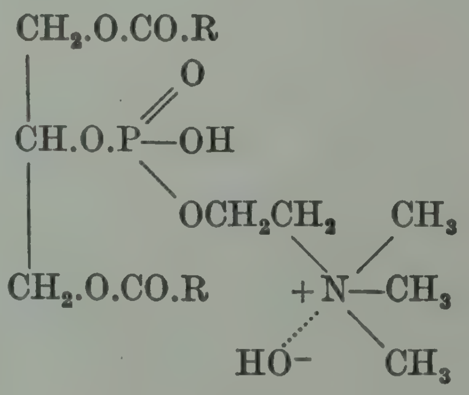

HOCH₂CH₂
CH₃
+N—CH₃
HO⁻
CH₃

choline
(trimethyl hydroxy ethyl ammonium hydroxide)

These various constituents are combined in the lecithins in the following way, where R is a fatty acid residue:


CH₂.O.CO.R
CH.O.CO.R
CH₂.O.P—OH
O
OCH₂CH₂
CH₃
+N—CH₃
HO⁻
CH₃

α-lecithin

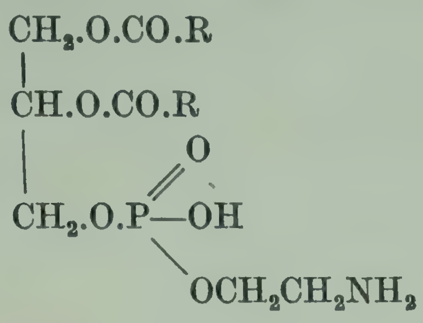

CH₂.O.CO.R
CH.O.P—OH
O
OCH₂CH₂
CH₃
CH₂.O.CO.R
+N—CH₃
HO⁻
CH₃

β-lecithin

The phosphoric acid grouping may be attached to either the α or the β carbon atom in the glycerol to form α-lecithin or β-lecithin. Lecithins, like the other phosphatides, are yellowish, greasy solids, soluble in all the fat solvents except acetone. This property enables them to be easily distinguished from the fats. On exposure to air they rapidly darken in colour and absorb water, forming a dark, greasy mass.

The lecithins can be attacked by the enzyme, lecithinase A, which removes one of the fatty acid residues, leaving a product known as *lysolecithin*, which has the property of causing hæmolysis of red blood cells. Lecithinase A is present in the venoms of many snakes and poisonous insects.

(b) *Kephalins* have properties similar to those of lecithins, but differ in structure by containing, instead of choline, the substance aminoethyl alcohol (ethanolamine, colamine), $H_2N—CH_2CH_2OH$:

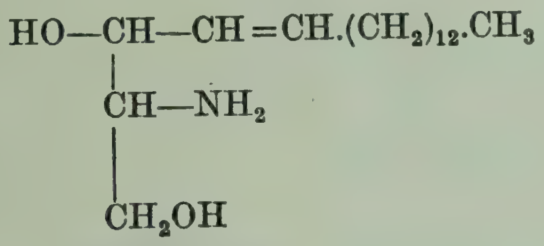


α-kephalin

Like lecithin they can lose one fatty acid residue with the production of a lysolipin, in this case, lysokephalin.

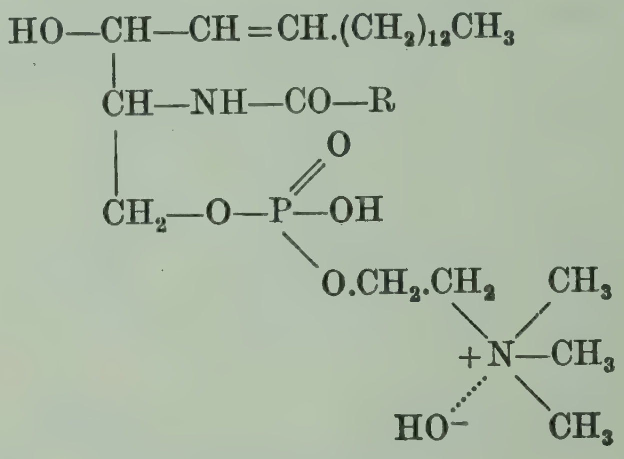
A second type of kephalin contains the amino-acid serine in place of ethanolamine and is known as *phosphatidyl serine.*

(c) *Sphingomyelins* are more complicated phosphatides which contain instead of glycerol the base sphingosine. On hydrolysis they yield fatty acids, phosphoric acid, choline and sphingosine.



$$\begin{array}{l} \text{HO—CH—CH=CH.(CH}_2)_{12}.\text{CH}_3 \\ \qquad | \\ \qquad \text{CH—NH}_2 \\ \qquad | \\ \qquad \text{CH}_2\text{OH} \end{array}$$

sphingosine



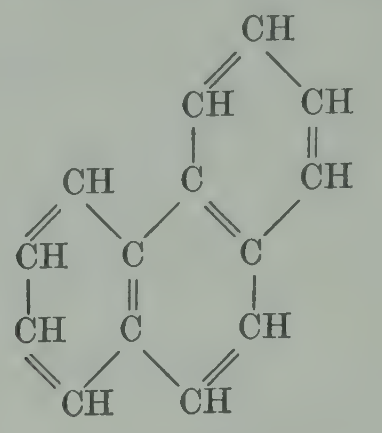
sphingomyelin

(d) *Lipositols* are plant phospholipids containing inositol (p. 13) esterified with fatty acids and with phosphoric acid which is usually also combined with a nitrogenous base.

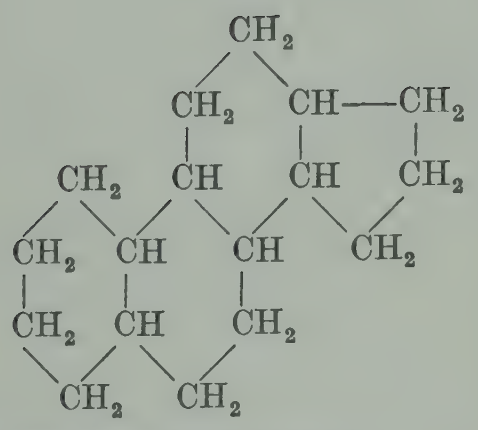
2. **Cerebrosides.**—The cerebrosides occur along with the phosphatides in particularly large amounts in brain tissue and in the myelin sheath of nerves. They include the substances *phrenosin* and *kerasin*, and are peculiar in containing the sugar, galactose, in their structure. On hydrolysis they yield fatty acids, sphingosine and galactose.

STEROIDS

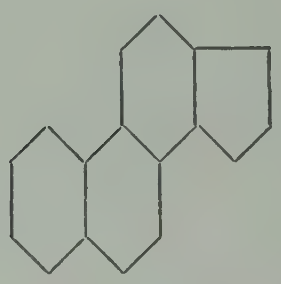
The steroids form a group of substances which are conveniently considered along with the lipids, although they differ from them greatly in chemical structure. Like the lipids they are soluble in the fat solvents and, in general, insoluble in water. As they are not hydrolysed by treatment with sodium hydroxide they form part of what is termed the *unsaponifiable fraction* of the lipid portion of tissues. The steroids include such substances as cholesterol and other sterols, the related D group of vitamins, the bile acids, the male and female sex hormones, the hormones of the adrenal cortex, and a number of miscellaneous substances, including toad poisons, cardiac glycosides of the digitalis group, and some of the carcinogenic hydrocarbons. The steroids are of great physiological importance, and although they show very varied biological activity, they all have one feature in common in that they contain the nucleus known as the perhydrocyclopenteno-phenanthrene ring system.



phenanthrene

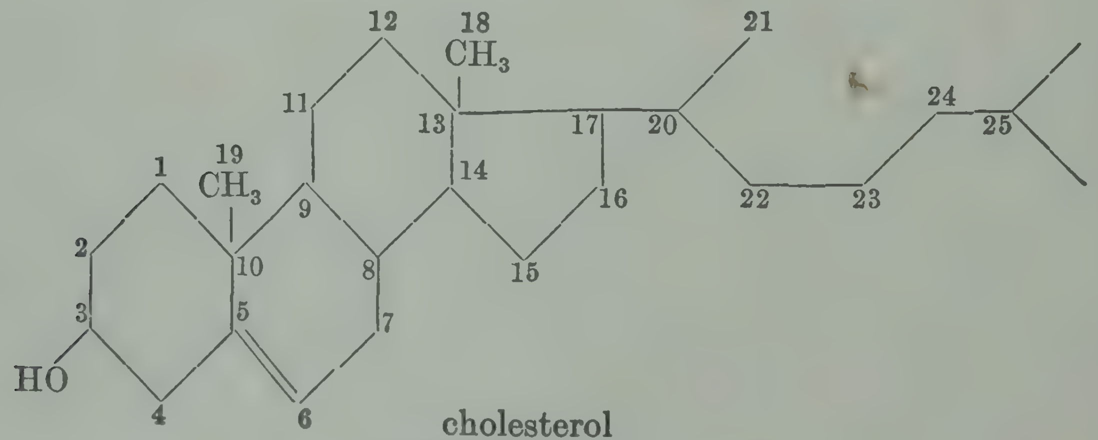


perhydrocyclopenteno-phenanthrene
(full formula)

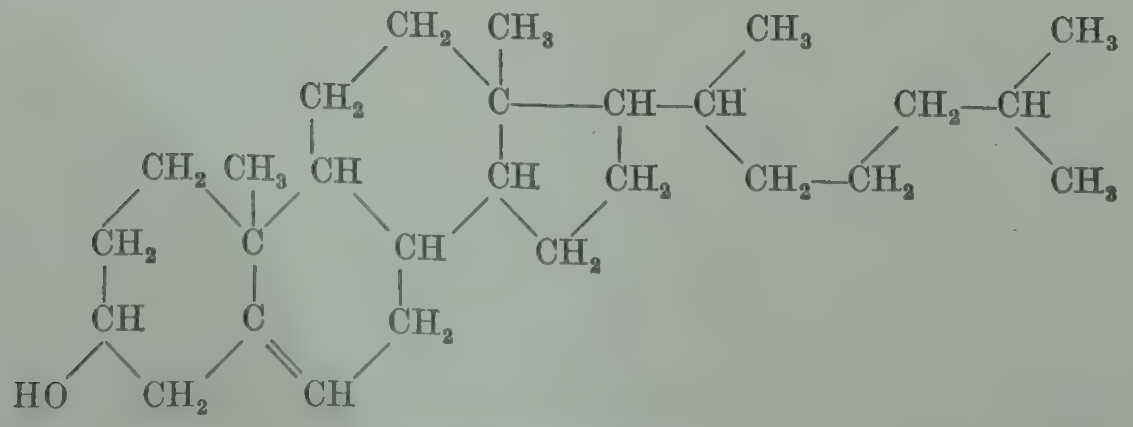


perhydrocyclopenteno-phenanthrene
(skeleton formula)

The Sterols. — The *sterols* are steroid alcohols. The best known is *cholesterol*, which is widely distributed in the tissues and may be conveniently isolated from gall-stones or brain tissue. It is particularly



cholesterol
(skeleton structure)



cholesterol (full structure)

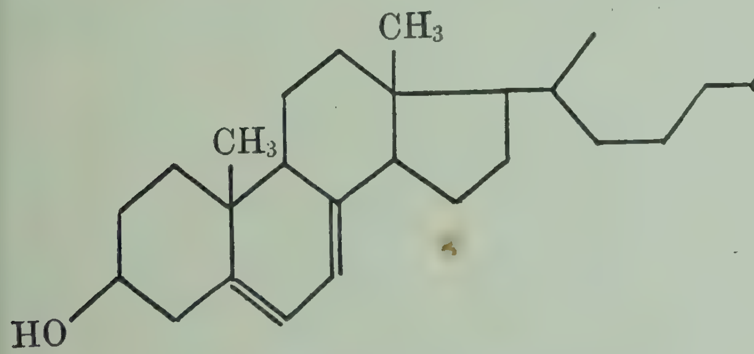
abundant in brain and nervous tissue, in the adrenal glands and in the skin, and is found also in egg-yolk. The pure substance is a white solid forming crystals of characteristic shape, insoluble in water but soluble in the fat solvents. It gives characteristic colour tests, of which the best known is probably the *Liebermann-Burchard test.* This consists of adding sulphuric acid and acetic anhydride to a solution of cholesterol in chloroform, when a blue colour is obtained. Cholesterol in chloroform solution also gives a reddish or purple colour when concentrated sulphuric acid is added. This is known as *Salkowski's test.* Both tests are given by some other sterols as well as by cholesterol.

Cholesterol is a derivative of the parent hydrocarbon cholestane (Table 3, 1).

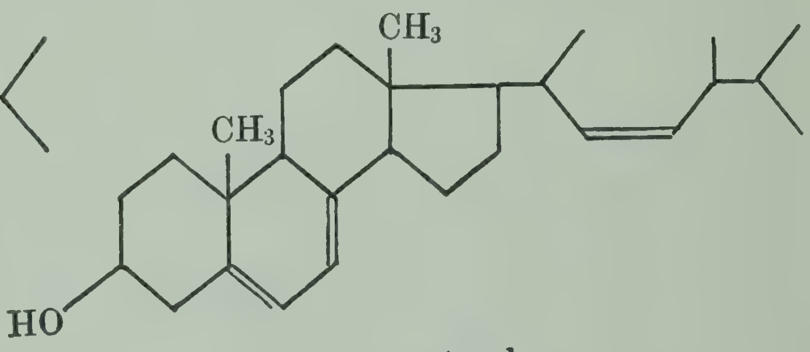
It will be noticed that cholesterol contains the cyclopenteno-phenanthrene ring system, with all the rings in the reduced state. A double bond is present between carbon atoms 5 and 6. Like all sterols, cholesterol has methyl groups attached to carbon atoms 10 and 13, and a hydroxyl group at position 3. All sterols, including cholesterol, have a long side chain at position 17, but its complexity varies from one sterol to another.

By virtue of its hydroxyl group cholesterol is a secondary alcohol and can therefore form esters. In the tissues it exists partly in the free state and partly in the form of esters which belong to the group of waxes (p. 21).

Many other sterols are known but only a few are of importance in the body. 7-*Dehydrocholesterol* occurs in skin and is converted by ultraviolet light to one form of vitamin D (p. 143). The plant sterol, *ergosterol,* which is found in yeast, likewise acts as the precursor of another form of vitamin D. *Coprosterol* is found in the fæces. Among the many plant sterols may be mentioned *sitosterol* and *stigmasterol.*



7-dehydrocholesterol



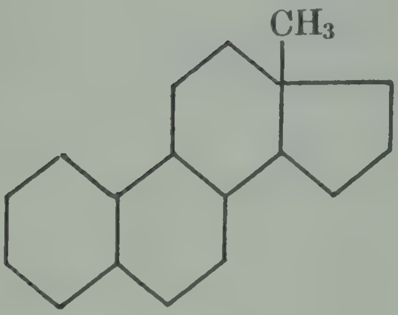
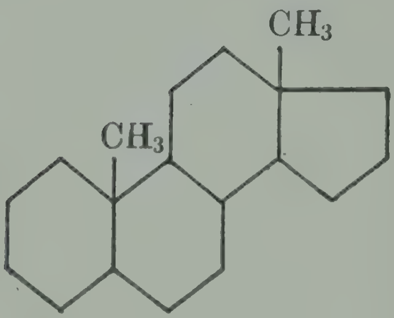
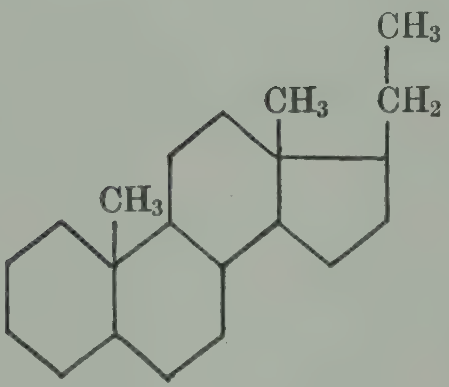
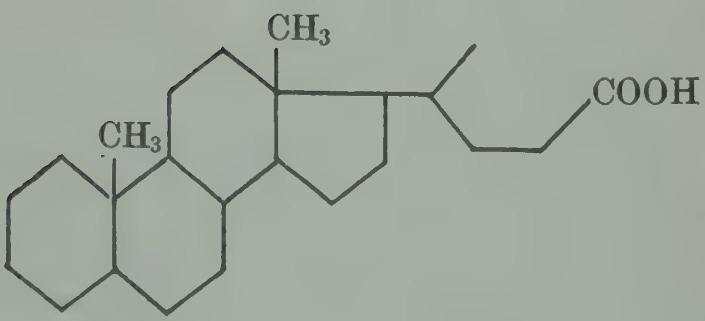
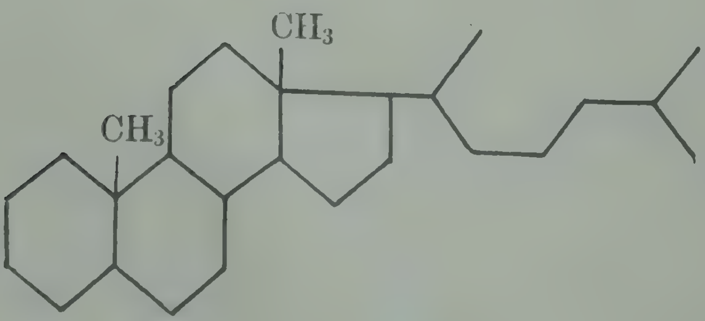
ergosterol

Nomenclature of Steroids.—Steroids may be conveniently classified according to the number of carbon atoms present in the ring structures plus side chains. For this purpose parent saturated hydrocarbons are taken. These are shown in Table 3, 1. Unsaturated compounds are indicated by the suffix *-ene* added to the root name of the hydrocarbon, and the position of the double bond is indicated by the symbol Δ and a superscript. The presence of an alcohol group is indicated by *-ol* and that of a keto group by *-one.* These groups are preceded by their position and configuration. Cholesterol can be described as Δ^5-*cholestene*-3-*ol.*

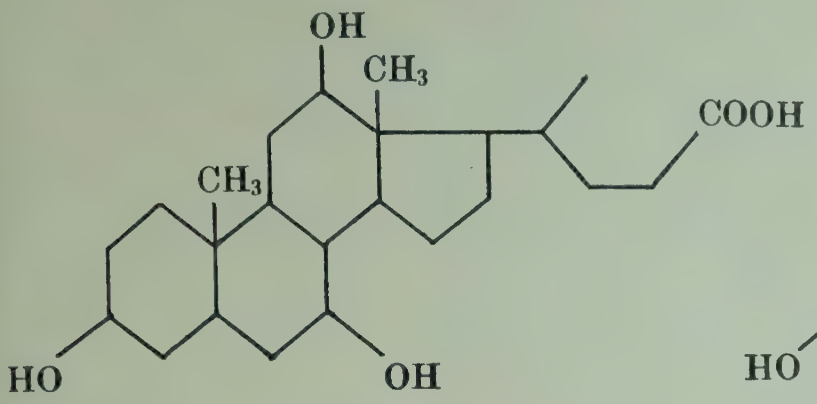
The Bile Acids.—The bile contains a number of steroid acids which are derivatives of the parent substance *cholanic acid.* They include *cholic acid, deoxycholic acid* and *lithocholic acid,* all of which contain the cyclopentenophenanthrene ring system, with a five-membered side chain on position 17 and with hydroxyl groups attached at other positions on the nucleus.

Table 3, 1

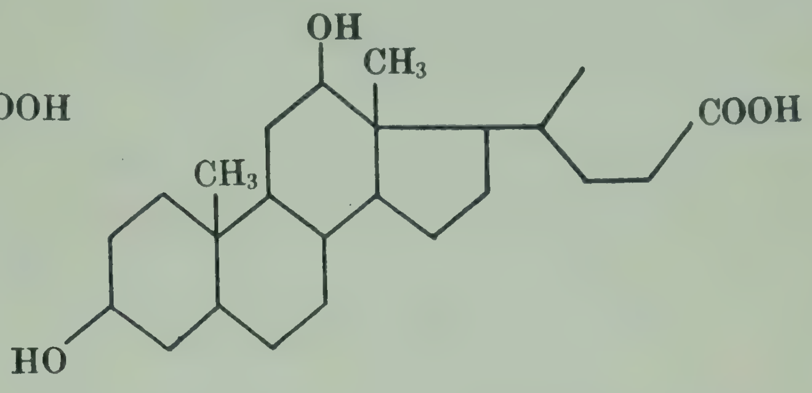
THE STEROIDS

No. of carbon atoms	Parent hydrocarbon (or acid)		Biologically important compounds
18		Œstrane	Œstrogens
19		Androstane	Androgens
21		Pregnane	Progestens Adrenal cortical hormones
24		Cholanic acid	Bile acids
27		Cholestane	Cholesterol

These acids are found in the bile in combination with glycine and taurine. They are discussed on p. 255.

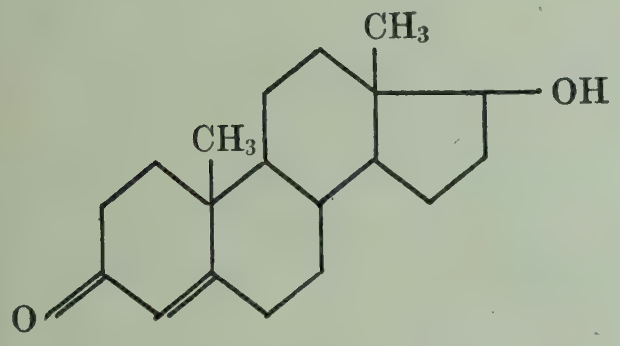


cholic acid

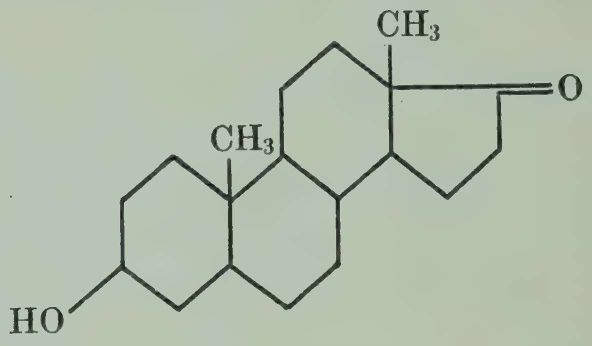


deoxycholic acid

The Sex Hormones.—The *male sex hormones* or *androgens* are discussed in Chapter 52. They are all derivatives of the parent hydrocarbon androstane, and include testosterone and androsterone.



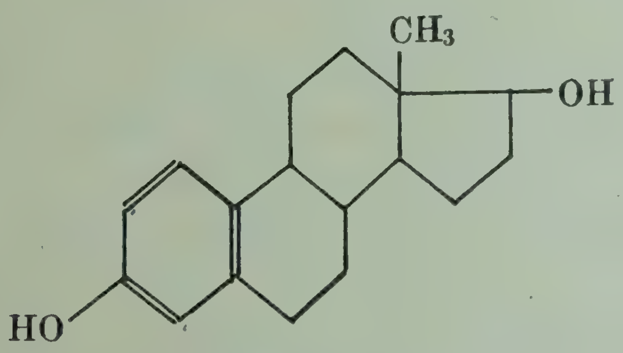
testosterone
(Δ^4-androstene-17-ol-3-one)



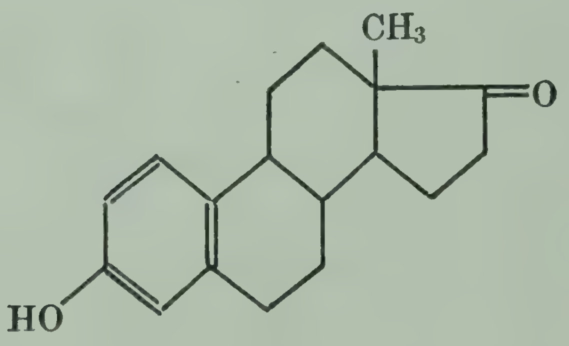
androsterone
(androstane-3-ol-17-one)

These androgens have no side chain in position 17 but have keto or hydroxyl groupings in positions 3 and 17. Those with keto groups in position 17 are known as 17-ketosteroids (17-oxosteroids). Their significance is discussed in Chapter 52.

The *female sex hormones* are of two groups, the *œstrogens*, which are the hormones characteristic of the ovary, and the *progestens* derived from the corpus luteum. The natural *œstrogens* are derivatives of the parent hydrocarbon *œstrane*, and include *œstradiol* and *œstrone*.



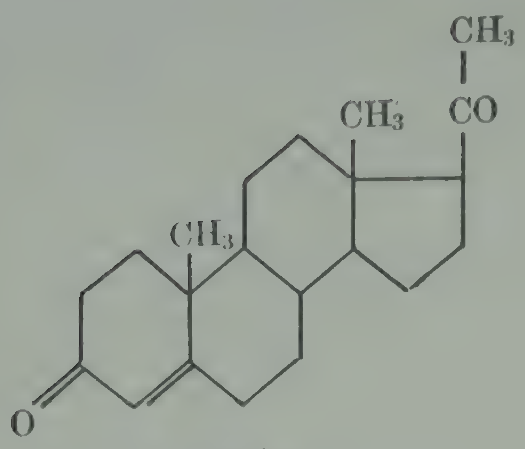
œstradiol
($\Delta^{1,3,5}$ œstratriene-3:17-diol)



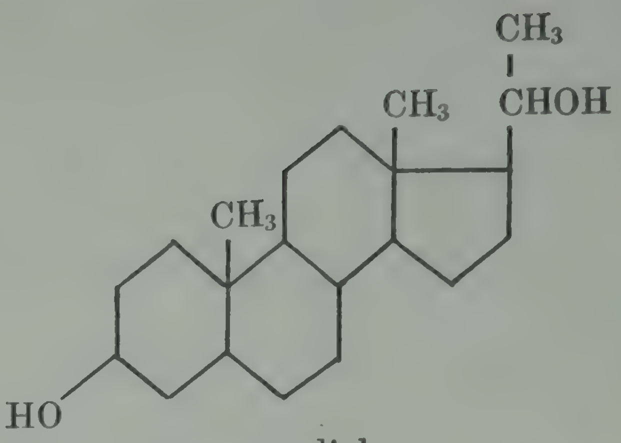
œstrone
($\Delta^{1,3,5}$ œstratriene-3-ol-17-one)

In œstradiol and œstrone the first ring is of the benzene type, containing three double bonds, a methyl group is attached to position 13 but not to position 10, there is no side chain, and carbon atoms 3 and 17 carry either hydroxyl or keto groupings.

The progesten *progesterone* is a derivative of the parent hydrocarbon *pregnane*.



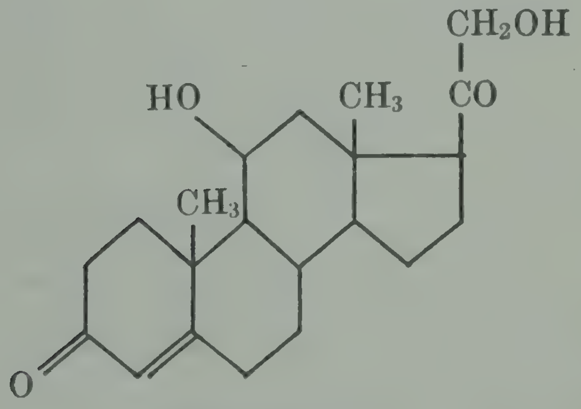
progesterone
(Δ^4 pregnene-3:20-dione)



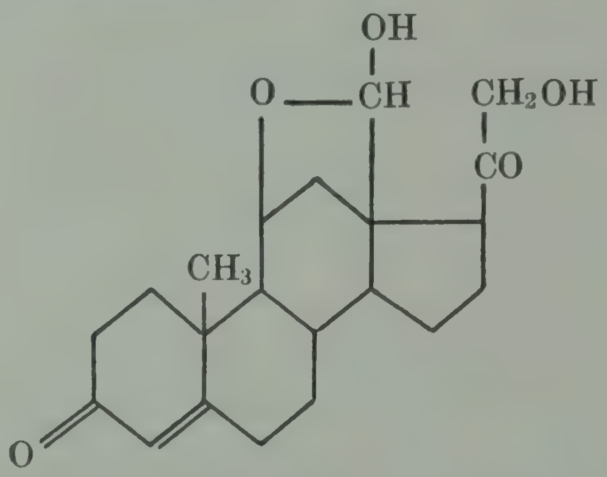
pregnanediol
(pregnane-3:20-diol)

It has a short side chain of two carbon atoms in position 17. A related compound is *pregnanediol*. The œstrogens and progesterone are discussed in greater detail in Chapter 52.

The *hormones of the adrenal cortex* also contain a two-carbon side chain in position 17. They include such compounds as *corticosterone* and *aldosterone* (*electrocortin*), discussed in Chapter 51.



corticosterone



aldosterone

Since all these steroids have very similar structures it has been suggested that they may all be derived from the same parent substance, such as cholesterol, although there is no direct evidence for this view. For example, a shortening of the side chain of cholesterol would give a substance similar to the bile acids, and further shortening of the side chain to two carbon atoms would give compounds similar to progesterone and corticosterone. Further oxidation with the removal of the side chain would yield compounds similar to the œstrogens or androgens.

Stereochemistry of the Steroids.—The steroids contain several asymmetric carbon atoms, for example at positions 3, 5 and 10, and can therefore exist in different stereoisomeric forms. The precise configuration at each asymmetric carbon atom is of great importance, but the details of this subject are beyond the scope of this book.

REFERENCES

Deuel, H. J. (1951). *The Lipids.* Vol. 1. London: Interscience.

Fieser, L. F. & Fieser, M. (1949). *Natural Products related to Phenanthrene*, 3rd ed. New York: Reinhold.

Haslewood, G. A. D. (1955). Recent developments in our knowledge of bile salts. *Physiol. Rev.* **35,** 178–199.

Holman, R. T., Lundberg, W. O. & Malkin, T. (1952). *The Chemistry of Fats and other Lipids.* London: Pergamon Press.

Lovern, J. A. (1955). *The Chemistry of Lipids of Biochemical Significance.* London: Methuen.

Wittcoff, H. (1951). *The Phosphatides.* London: Chapman and Hall.

CHAPTER 4

SOME PHYSICO-CHEMICAL CONSIDERATIONS

At this stage in the discussion of the components of the body it is desirable to consider briefly some physico-chemical factors which influence the behaviour of biological systems. It is assumed that the reader is familiar with the elementary aspects of physical chemistry, but a few salient features will be covered in this chapter by way of revision.

THE STRENGTH OF SOLUTIONS

Although it has long been customary to express the strength of solutions as per cent. by volume, that is as grams of solute per 100 ml. solution, it is much more useful to think of concentration in terms of gram molecules per litre.

A molar solution (M-solution) contains one gram-molecular weight (one mole) per litre of solution. A millimolar solution (mM) contains 0·001 gram-molecular weight (one millimole) per l. A normal solution (N-solution) contains one gram equivalent per litre, the equivalent being that amount of a substance which is capable of reacting with or releasing, directly or indirectly, 8 g. oxygen, or 1·008 g. hydrogen.

One milliequivalent (m.equiv) is 0·001 equivalent. For example the molecular weight of HCl is 36·5. A molar solution (M-HCl) contains 36·5 g./l. and a millimolar solution contains 36·5 mg./l. Since the equivalent weight of HCl is the same as the gram-molecular weight, a normal solution (N-HCl) contains one equivalent/l. or 36·5 g./l. while 0·001 N-HCl contains 36·5 mg./l.

The molecular weight of H_2SO_4 is 98. A molar solution (M-H_2SO_4) contains 98 g./l. and a millimolar solution contains 98 mg./l. Since the sulphate ion is divalent the equivalent weight of H_2SO_4 is 49, and hence a normal solution (N-H_2SO_4) contains 49 g./l. while 0·001 N-H_2SO_4 contains 49 mg./l.

It is customary nowadays to express the concentration of solutes (especially inorganic ions) in biological fluids in terms of millimoles per litre or of milliequivalents (m.equiv) per litre. The term milliequivalent is used for this purpose with reference both to molecules and to ions.

In sodium chloride one atom of sodium is combined with one of chlorine. A milliequivalent of sodium chloride is therefore 23 + 35·5, that is 58·5 mg. Since sodium chloride dissociates completely in dilute solutions into sodium and chloride ions, a solution containing 58·5 mg./l. (1 mM-NaCl) contains one milliequivalent of sodium ions per litre (1 m.equiv Na^+/l.) and one milliequivalent of chloride ions per litre (1 m.equiv Cl^-/l.).

The expression milliosmole is used in connexion with osmotic pressure (see below). The osmotic pressure of a solution is proportional to the number of particles in solution. Hence for a non-electrolyte such as glucose 1 milliosmole = 1 millimole. In the case of an electrolyte such as sodium chloride which dissociates completely in dilute solution into Na^+ ions and Cl^- ions, 2 milliosmoles ≡ 1 millimole. For $CaCl_2$ 3 milliosmoles ≡ 1 millimole.

To take another example, the concentration of sodium (as Na^+ ion) in blood plasma is 330 mg./100 ml. The atomic weight of Na is 23, and since the Na^+ ion is monovalent, one milliequivalent is 23 mg. Hence the concentration of Na^+ in plasma is 330/23 m.equiv/100 ml., i.e. 3300/23 m.equiv/l. or 143·4 m.equiv/l. (which is the same as 143·4 millimoles/l. or 143·4 milliosmoles/l.).

Similarly the concentration of calcium (as Ca^{++} ion) in blood plasma is 10 mg./100 ml. The atomic weight of Ca is 40, and since the Ca^{++} ion is bivalent, one milliequivalent is 20 mg. Hence the concentration of Ca^{++} in plasma is 10/20 m.equiv/100 ml., i.e. 100/20 m.equiv/l. or 5 m.equiv/l. (which is the same as 2·5 millimoles/l. or 2·5 milliosmoles/l.).

One advantage of this method of expression is that it enables us to equate anions with cations as shown in the following example:—

AVERAGE "ACID-BASE" COMPOSITION OF VENOUS PLASMA

Cations	mg./100 ml.	m.equiv/l.	Anions	mg./100 ml.	m.equiv/l.
Na^+	330	143·4	HCO_3^-	60 ml. of CO_2	27·0
K^+	20	5·1	Cl^-	365	103·0
Ca^{++}	10	5·0	$HPO_4^{=}$	5	3·0
Mg^{++}	3	2·5	$SO_4^{=}$	1	1·0
			Proteins	5090	16·0
			Organic acids	—	6·0
Totals	..	156·0	..	..	156·0

This table makes it clear that the cations are exactly balanced by the anions and shows that sodium, though present in smaller weight than chloride, can balance all the chloride bicarbonate and other acid radicals as well as some of the protein.

One millimole of a gas occupies 22·4 ml. at s.t.p. Hence 60 ml. CO_2/100 ml. corresponds to 60/22·4 or 2·7 millimoles/100 ml. or 27 m.equiv/l.

HYDROGEN-ION CONCENTRATION

Water is ordinarily split up to a very slight extent into hydrogen ions * and hydroxyl ions:

$$H_2O \rightleftharpoons H^+ + OH^-.$$

The dissociation is so slight that the concentration of hydrogen ions is exceedingly small. It can be shown by measurements of electrical conductivity and by other methods that the concentration of hydrogen ions $[H^+]$ in pure water at 22° is only one ten-millionth of a gram per litre, that is $[H^+]=10^{-7}$ g. ions per litre, or 1 g. H^+ in 10,000,000 litres. Since equal numbers of hydrogen and hydroxyl ions are produced by the ionization of pure water the concentration of hydroxyl ions in water at 22° is also 10^{-7} g. ions per litre. The dissociation of water is a reversible reaction. At equilibrium the rates of the forward and reverse reactions are equal and the equilibrium constant K is given by the equation:

$$K=\frac{[H^+]\,[OH^-]}{[H_2O]}$$

---

* Although the hydrogen ion is now regarded by physical chemists as being hydrated (H_3O^+), the argument which follows on this page still holds in all essentials.

where $[H^+]$ is the concentration of hydrogen ions.
$[OH^-]$,, ,, ,, hydroxyl ions.
$[H_2O]$,, ,, ,, undissociated water molecules.

Since $[H^+]$ and $[OH^-]$ are very small in comparison with $[H_2O]$, any increase or decrease in the ionization of the water, that is the formation of hydroxyl and hydrogen ions from water molecules, has an insignificant effect on $[H_2O]$ which may be regarded as a constant,

that is $$\frac{[H^+]\,[OH^-]}{\text{constant}}=K$$

or $$[H^+]\,[OH^-]=K_w \text{ (another constant).}$$

As we have already seen in pure water at 22° $[H^+]=10^{-7}$ and $[OH^-]=10^{-7}$ so that $K_w=10^{-7}\times10^{-7}=10^{-14}$. This relationship also holds when salts, acids, or alkalis are dissolved in the water. If acid is added, the concentration of hydrogen ions, $[H^+]$, is increased but, since the product of $[H^+]$ and $[OH^-]$ is constant at 10^{-14}, it follows that $[OH^-]$ must decrease. Similarly if alkali is added $[OH^-]$ is increased, and this increase must be accompanied by a corresponding decrease in $[H^+]$. For instance if by adding alkali $[OH^-]$ is increased from 10^{-7} to 10^{-5} then $[H^+]$ must decrease from 10^{-7} to 10^{-9}.

When $[H^+]=[OH^-]$ the solution is said to be neutral. If $[H^+]$ is greater than $[OH^-]$ the solution is acid. If $[H^+]$ is less than $[OH^-]$ the solution is alkaline.

To express hydrogen-ion concentration in convenient units, the pH notation was devised by Sörensen. The pH of a solution is defined as the negative value of the logarithm to the base 10 of the hydrogen-ion concentration, that is $pH=-\log_{10}[H^+]$. Thus if $[H^+]=10^{-7}$ the $pH=-\log_{10}10^{-7}=7$, or if $[H^+]=10^{-2}$, $pH=-\log_{10}10^{-2}=2$, and so on. It follows that a neutral solution has pH 7, an acid solution has $pH<7$, and an alkaline solution has $pH>7$. It should be noted that a small figure indicates a high acidity and conversely a larger figure indicates a low acidity; a solution of pH 2 is much more acid than one of pH 6.

Let us take the case of 0·1 N HCl. If it is completely dissociated the solution is 0·1 N with respect to hydrogen ions, that is, $[H^+]=10^{-1}$ and $pH=-\log 10^{-1}=1$. Or if we have 0·1 N NaOH then $[OH^-]=10^{-1}$ and since $[H^+]\times[OH^-]=10^{-14}$ it follows that $[H^+]=10^{-13}$ and $pH=13$.

On the other hand, if $[H^+]=2{\cdot}0\times10^{-5}$,

$$\begin{aligned}\text{by definition pH} &= -\log[H^+]\\ &= -\log(2{\cdot}0\times10^{-5})\\ &= -\log 2-\log 10^{-5}\\ &= -0{\cdot}3010+5\\ &=4{\cdot}7.\end{aligned}$$

The converse calculation is carried out as follows.

If we have a solution of pH 7·4 then, from the definition of pH,

$$\begin{aligned}-\log[H^+] &= pH=7{\cdot}4\\ &=8{\cdot}0-0{\cdot}6\\ &=\log 10^8-\log 3{\cdot}98=-(\log 3{\cdot}98-\log 10^8)\\ &=-(\log 3{\cdot}98+\log 10^{-8})\\ -\log[H^+] &= -\log(3{\cdot}98\times10^{-8})\\ [H^+] &= 3{\cdot}98\times10^{-8}\text{ g. ions per litre.}\end{aligned}$$

Similarly it can be shown that at pH 7·3 $[H^+] = 5{\cdot}01 \times 10^{-8}$ g. ions per litre.

Hence an apparently small increase in the pH of blood, for example from 7·4 to 7·3, represents an actual increase in $[H^+]$ of 3·98 to $5{\cdot}01 \times 10^{-8}$ g. ions per litre, that is an increase of 26 per cent.

The pH of a solution is measured either by indicators or by electrical methods. For a description of these procedures a text-book of practical biochemistry should be consulted.

BUFFERS

A strong acid is one which dissociates freely in aqueous solution. Hydrochloric acid dissolved in water dissociates almost completely into hydrogen ions and chloride ions,

$$HCl \rightleftharpoons H^+ + Cl^-.$$

On the other hand, a weak acid is one which is poorly ionized in solution. If acetic acid is dissolved in water, it dissociates only to a slight extent into hydrogen ions and acetate ions,

$$HA \rightleftharpoons H^+ + A^-.$$

When equilibrium is reached,

$$\frac{[H^+][A^-]}{[HA]} = K$$

$[H^+]$ = concn. of H ions,
$[A^-]$ = concn. of acetate ions,
$[HA]$ = concn. of undissociated acetic acid molecules,

and the equilibrium constant K is called in this case a *dissociation constant*. Each weak acid has its own characteristic dissociation constant.

Suppose sodium acetate is added to a solution of acetic acid. Since sodium acetate is almost completely ionized we are in effect adding a high concentration of acetate ions.

Now the equation

$$HA \rightleftharpoons H^+ + A^-$$

represents an equilibrium, and it follows from the law of mass action that addition of acetate ions will drive the reaction to the left and more undissociated acetic acid molecules will be formed, so that the dissociation of acetic acid is greatly diminished.

Therefore in the mixture of sodium acetate and acetic acid the sodium acetate is completely dissociated into sodium ions and acetate ions, while the acid is dissociated only to a very small extent.

The equation

$$\frac{[H^+][A^-]}{[HA]} = K \text{ (the dissociation constant)}$$

holds for the acid, whether salt is present or not, and if we take logarithms of both sides we get

$$\log K = \log \frac{[H^+][A^-]}{[HA]} = \log [H^+] + \log \frac{[A^-]}{[HA]}$$

$$-\log K = -\log [H^+] - \log \frac{[A^-]}{[HA]}$$

$$pK = pH - \log \frac{[A^-]}{[HA]} \quad \text{(where } pK = -\log K \text{ by analogy with Sörensen's definition of pH)}$$

$$pH = pK + \log \frac{[A^-]}{[HA]}.$$

Since the acetate ions are nearly all from the salt which is fully dissociated, it follows that

$$[A^-] \equiv [\text{salt}].$$

Moreover [HA] represents the total concentration of acid, since hardly any of the weak acid is ionized.

Hence $$[HA] \equiv [\text{acid}]$$

and $$\text{pH} = \text{pK} + \log \frac{[\text{salt}]}{[\text{acid}]}.$$

This is an equation of the greatest importance.

If the salt concentration and the acid concentrations are equal,

$$[\text{salt}] = [\text{acid}] \quad \text{then} \quad \frac{[\text{salt}]}{[\text{acid}]} = 1 \text{ and } \log \frac{[\text{salt}]}{[\text{acid}]} = 0.$$

Hence pH = pK.

Therefore when the acid is half neutralized, pH = pK. If we take 100 ml. 0·1 N acid and add 1 ml. of 0·1 N alkali, and if the pK of the acid is 5,

$$\begin{aligned} \text{pH} &= \text{pK} + \log \frac{[\text{salt}]}{[\text{acid}]} \\ &= 5 + \log \tfrac{1}{99} \\ &= 5 - \log 99 \\ &= 5 - 1{\cdot}996 \\ &= 3{\cdot}004. \end{aligned}$$

The pH of the solution depends on the ratio of the salt to the acid and not on the absolute quantity of either. This is one way by which we can calculate the pH of a mixture of a salt and an acid.

A mixture of a weak acid and its salt, say acetic acid and sodium acetate, has a definite pH determined by the ratio of the salt and acid concentrations as in the above equations. If now a strong acid, such as hydrochloric acid, is added, the following reaction takes place,

$$HCl + NaA = NaCl + HA.$$

Since the sodium chloride is neutral and the acetic acid is feebly ionized in the presence of sodium acetate, the end-result is that the concentration of hydrogen ions is almost unaltered; in other words, the pH of the solution hardly changes at all. If we had added hydrochloric acid to water there would have been a great rise in hydrogen ions and the pH would have fallen very considerably.

Such mixtures of weak acids and their salts are called *buffers*. A buffer solution is one to which we can add acid or alkali without much change in pH. They minimize the effect of adding acid and alkali and prevent large changes in hydrogen-ion concentration. We have seen that acetate: acetic acid mixtures act as buffers. Phosphate buffers occur in the body and are commonly used in the laboratory. The addition of hydrochloric acid to a solution of disodium hydrogen phosphate results in the following reaction:

$$Na_2HPO_4 + HCl = NaH_2PO_4 + NaCl$$

in which Na_2HPO_4 is very slightly alkaline, NaCl is neutral, NaH_2PO_4 is very slightly acid. The pH change is therefore small, or in other words the addition of a strong acid to an almost neutral solution results in an almost neutral solution.

THE COLLOIDAL STATE

The word 'colloid' was originally used by Thomas Graham to indicate those substances which in aqueous solutions fail to pass through parchment membranes. Substances in solution which diffuse readily through such membranes were named 'crystalloids.' Although it was originally assumed that colloids were substances of such high molecular weight that they were unable to pass through the pores in the membranes, it is now known that under suitable conditions many materials usually considered to be crystalloids can be made to behave as colloids. Consequently it is usual nowadays to apply the term 'the colloidal state' to particles varying in diameter between 1 and 500 millimicrons, which may be either single large molecules such as protein molecules or aggregates of smaller molecules.

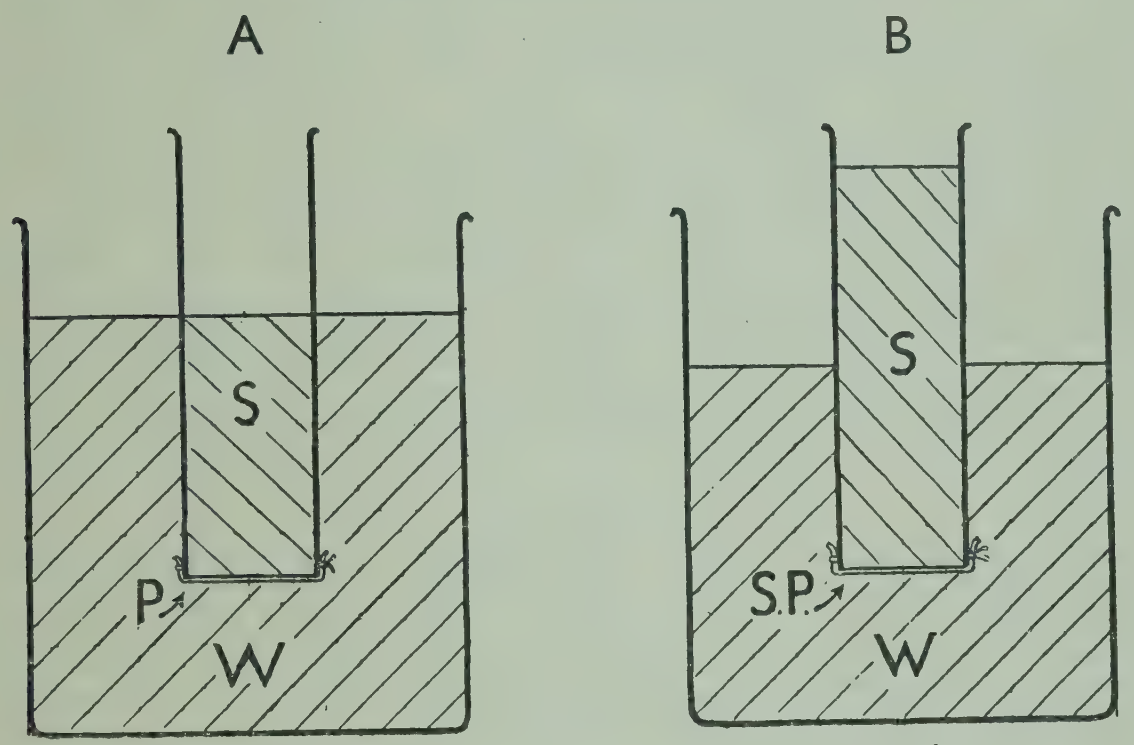
Colloidal particles are electrically charged, and as the charges on the individual particles are like, neighbouring particles repel one another so that they are maintained in suspension. If the charges are neutralized, the colloidal suspension becomes unstable and the colloid particles may gradually settle out of solution.

Some colloids known as *hydrophobic colloids* or *suspensoids* are not hydrated, and are readily precipitated when the charge on their particles is neutralized by the addition of an electrolyte. A simple example is a gold sol. On the other hand, most biologically important colloids are *hydrophilic colloids* or *emulsoids*, in which the particles are surrounded by a shell of water molecules which acts as a protective coating making them less susceptible to precipitation by electrolytes. They can, however, be precipitated by alcohol which dehydrates them or by strong solutions of salts. This is the basis of the well-known procedure of 'salting out' proteins from aqueous solutions.

The most important colloids in biological material are the proteins. They exist in aqueous solution as large, heavily hydrated molecules which may be separated readily from simple salts by the process of *dialysis* discussed below.

OSMOTIC PRESSURE

If a vessel, S, containing a salt solution is separated from one containing pure water, W, by a collodion membrane (Fig. 4, 4, A), salt diffuses through the membrane until its concentration on both sides, that is in vessels S and W, is the same. This is an example of simple *diffusion*. If, however, the membrane, while freely permeable to water molecules, is impermeable to the solute molecules, that is if the membrane is *semi-permeable*, the result is quite different (Fig. 4, 4, B). On the water side, the membrane is being continually bombarded by water molecules, which pass through; on the other side the membrane is being bombarded both by solute molecules (which do not pass through) and by water molecules (which *do* pass through), so that the number of water molecules able to reach the membrane and pass through from this side is smaller than the number able to reach the other side of the membrane and pass in the reverse direction. Hence water passes from W into S and the concentration of the solution in S tends to fall. The level of liquid in S therefore tends to rise until the hydrostatic pressure of fluid in S exactly counterbalances the attractive power of the solution in S for water molecules. This attractive power is usually termed the *osmotic pressure*, that is the pressure which would have to be applied to S so as just to prevent the entrance of water molecules. It should be noted that the osmotic pressure of a solution can only be demonstrated by using a semipermeable membrane. It is not easy to make artificial membranes to show the osmotic pressure of salts; it is easier to demonstrate the phenomenon with substances of large molecular size, such as proteins or polysaccharides, which cannot pass through easily prepared membranes such as cellophane. To demonstrate osmotic phenomena it is not necessary that the fluid in W of Fig. 4, 4, B, should be pure water. If two solutions of different concentrations are separated by a semipermeable membrane, water tends to pass from the more dilute to the more concentrated until both are at the same osmotic pressure.



Simple diffusion. Osmosis.

FIG. 4, 4

S, salt solution.
W, water.
P, permeable membrane.
SP, semi-permeable membrane.

The osmotic pressure of a solution obviously depends upon the number of solute molecules which compete with water molecules in striking the surface of the membrane in S. It can in fact be shown that the osmotic pressure of a solution containing 1 g. molecule of solute in 22·4 l. of water is one atmosphere. The osmotic pressure of a solution therefore depends on its molecular concentration. Molar solutions of non-ionizable solutes all have the same osmotic pressure, although their concentrations expressed as g./litre are very different. Thus a molar solution of glucose (180 g./litre) has the same osmotic pressure as a molar solution of sucrose (342 g./litre), namely 22·4 atmospheres, but a 1 per cent. solution of glucose has an osmotic pressure of 946 mm Hg, which is nearly twice the osmotic pressure of a 1 per cent. solution of sucrose (498 mm Hg). In the same way a 1 per cent. solution of a substance of very high molecular weight such as albumin (68,000) has a very low osmotic pressure, namely 2·5 mm Hg.

If a solute ionizes in water the osmotic pressure depends on the number of ions rather than on the number of molecules. Thus if we assume that sodium chloride and calcium chloride are completely dissociated in aqueous solution, then M-NaCl will have twice the osmotic pressure, and M-$CaCl_2$ three times the osmotic pressure, of M-glucose. In other words, a litre of a molar solution of glucose, sodium chloride or calcium chloride respectively, contains one, two or three osmoles.

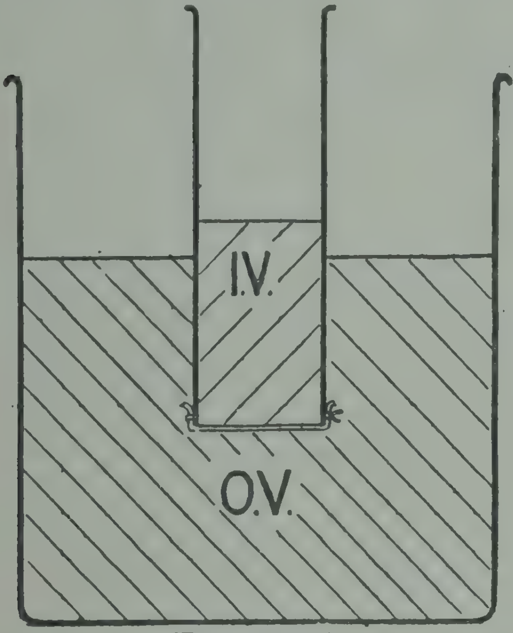


Fig. 4, 5

Dialysis

If a solution containing a simple salt, such as sodium chloride, and a protein is placed in an inner vessel such as I.V. (Fig. 4, 5), and is separated from pure water in the outer vessel O.V. by a *semipermeable* membrane which allows the passage of small, but not of large, molecules, the osmotic pressure of the solution in I.V. tends to cause water to pass from O.V. to I.V. At the same time salt passes out from I.V. to O.V., and if the water in O.V. is continually renewed the protein solution in I.V. eventually becomes salt-free. This is the process of *dialysis*.

A more complex situation, however, arises if a semipermeable membrane is used to separate a solution of sodium chloride (NaCl) and a solution of the sodium salt of some such material as congo red (NaA), where the ion A^- is not able to pass the membrane,

(1)	membrane	(2)
Na^+		Na^+
A^-		Cl^-

When equilibrium is reached chloride ions are found on both sides of the membrane by the transfer of Na^+ and Cl^- ions in equal amounts across the membrane thus,

(1)	membrane	(2)
Na^+		Na^+
Cl^-		Cl^-
A^-		

and it can be proved on thermodynamical grounds that, at equilibrium, the product of the concentrations of sodium and chloride ions on side (1) is equal to the product of the sodium and chloride concentrations on side (2), that is

$$[Na^+]_1\,[Cl^-]_1=[Na^+]_2\,[Cl^-]_2. \tag{1}$$

This is a very simple case of what is known as the *Donnan equilibrium*.

In order that the system may be electrically neutral it is necessary that

$$[Na^+]_2=[Cl^-]_2.$$

Substituting this in equation (1) we have

$$[Na^+]_1\,[Cl^-]_1=[Cl^-]_2^2. \qquad (2)$$

Electrical neutrality in side (1) is expressed by

$$[Na^+]_1=[Cl^-]_1+[A^-]$$
$$\therefore\quad [Na^+]_1 > [Cl^-]_1.$$

Multiply both sides by $[Cl^-]_1$

$$[Na^+]_1\,[Cl^-]_1 > [Cl^-]_1^2.$$

Substituting from equation (2) we have

$$[Cl^-]_2^2 > [Cl^-]_1^2$$
$$[Cl^-]_2 > [Cl^-]_1.$$

Thus the sodium chloride is more concentrated on the side of the membrane which is free from NaA.

Since the distribution of ions on the two sides of the membrane is unequal, a potential difference, the *membrane potential*, is set up between the two solutions separated by the membrane. It can be shown that the membrane potential E_{m} is given by the equation $E_{\mathrm{m}}=2{\cdot}303\dfrac{RT}{F}\log_{10}(1/\lambda)$, where R = the thermodynamic gas constant, T = the absolute temperature, F = the Faraday, and

$$\lambda=\frac{[Na^+]_2}{[Na^+]_1}=\frac{[Cl^-]_1}{[Cl^-]_2}.$$

Membrane potentials occur in many parts of the body, especially in nerve fibres and in muscle cells, where the concentration of potassium ions is much greater inside the cell than it is outside. The potential, measured experimentally, between the inside and outside of nerve cells and also muscle cells, agrees well with the value calculated from the formula just given (see Chapter 41).

Colloid Osmotic Pressure

Semi-permeable membranes are not uncommon in the body, for many body fluids contain both diffusible ions and non-diffusible substances such as proteins. A very simple example is to be found in the red blood cell. If human red blood cells are suspended in 0·5 per cent. sodium chloride solution which has a lower osmotic pressure than the cell contents (a *hypotonic* solution), water and salt pass into the cells until they swell and burst. If red blood cells on the other hand are suspended in 1·4 per cent. sodium chloride solution which has a higher osmotic pressure (a *hypertonic* solution), water and salt pass out of the red cells so that they shrivel up and become *crenated*. At some intermediate concentration of sodium chloride (about 0·9 per cent.) the cell neither swells nor shrinks, and at this concentration the sodium chloride solution is said to be *isotonic* with the cell contents.

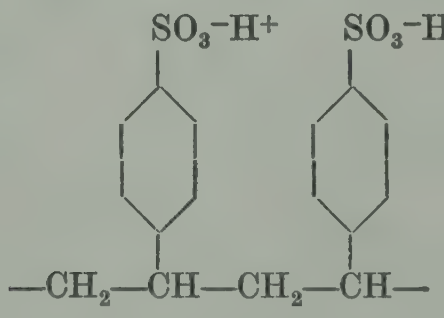
Blood plasma contains both readily diffusible, small, non-colloidal molecules or ions and non-diffusible, large colloidal particles. The total osmotic pressure of blood is equivalent to about 7 atmospheres, whereas the proportion due to the plasma proteins (the *colloid osmotic pressure*) is equivalent to only some 28 mm Hg. Nevertheless this latter proportion is of great physiological importance, since it determines the distribution of fluid between blood and tissue fluid. The difference in the osmotic pressure of blood and of lymph or tissue fluid is due entirely to plasma protein and is estimated to be about 22 mm Hg. This difference represents the effective osmotic pressure or *oncotic pressure* of plasma. As we shall see in Chapter 35, in association with the hydrostatic pressure in the capillaries, it controls the flow of water between blood and tissue fluid.

ION EXCHANGERS

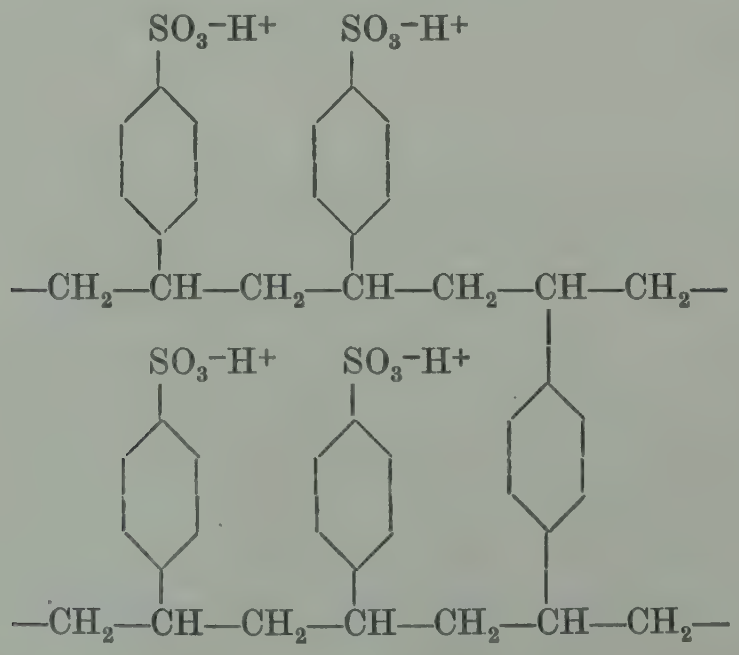
A technique recently introduced in biochemistry and medicine is the application of ion exchangers to biological problems. It has proved so valuable that a brief description of the theoretical background is introduced here.

Modern ion exchangers are usually synthetic resins insoluble in water, of very high molecular weight and of chain-like structure with the active groups projecting at intervals. Parallel chains may be joined by cross-linkages to give a lattice-like structure.

In the *cation exchangers* the active groups may be —OH, —COOH or —SO_3H. One segment of a chain of the third type may be represented thus:



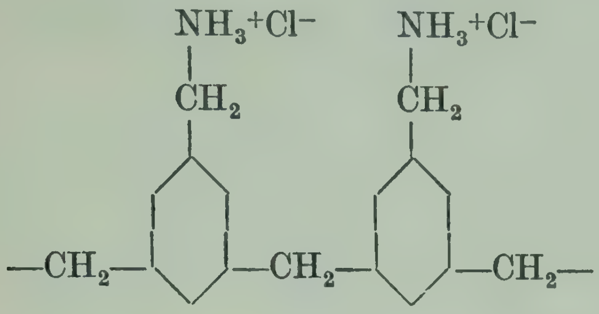
or where cross-linkages are present, thus:


$SO_3^-H^+$
$SO_3^-H^+$
—CH_2—CH—CH_2—CH—CH_2—CH—CH_2—
$SO_3^-H^+$
$SO_3^-H^+$
—CH_2—CH—CH_2—CH—CH_2—CH—CH_2—

The resin can exist both in the form of the free acid and of its salts, in which —$SO_3^-H^+$ becomes —$SO_3^-Na^+$ or —$SO_3^-K^+$ according to the cation involved. The use of such resins may be illustrated by a simple example. Suppose we wish to convert the potassium salt of penicillin to the sodium salt. We can do this by bringing a solution of the potassium salt into contact with a suitable cation exchanger in the sodium form. This solid gives up some of its sodium ions and takes potassium ions in exchange.

The process is reversible, but by passing the solution of potassium penicillin down a sufficiently long column of granules of ion exchanger we can ensure that the fluid which emerges contains only sodium penicillin.

Anion exchangers are usually aliphatic or aromatic bases containing amino groups thus:



The chloride form is shown in the diagram, but chloride ions may be exchanged for other anions such as sulphate.

Ion exchangers are extensively used in the softening of hard water in bulk by the replacement of calcium by sodium. They have been used in emergency kits on life rafts for making sea water potable. Each kit consists of a plastic bag containing six tablets of high capacity cation exchanger with silver ion as the replaceable cation. One tablet is put in the bag with a pint of sea water and the bag is kneaded to accelerate the reaction. The exchanger removes sodium and magnesium ions from the sea water, replacing them by silver ions which combine with the chloride ions of the sea water to form insoluble silver chloride which is strained off by canvas, leaving the water fit to drink.

Columns of ion exchangers are extensively used in biochemistry for the separation of amino-acids, nucleotides and other like substances, and they are also employed in clinical medicine. For example, a sulphonated resin in the potassium and ammonium form is used in cases of heart failure, nephrosis or hypertension where a sodium-free diet is desirable. The resin, when ingested, accepts sodium ions from the intestinal contents, replacing them by potassium and ammonium ions. This cationic exchange, a dynamic process, limits the availability of the sodium in the diet, thereby indirectly reducing the excessive sodium concentration of the body's extracellular fluid.

The resin is active throughout a large segment of the gastro-intestinal tract. In the stomach it liberates ammonium and potassium ions, and picks up hydrogen and some sodium ions. In the upper part of the small intestine, sodium, potassium and a small amount of calcium and magnesium ions are absorbed by the resin. In exchange, the resin releases ammonium, potassium and some hydrogen ions which were previously picked up from the stomach. Thus cationic interchange is constantly taking place. The ammonium ions which are released by the resin combine in part with chloride to form ammonium chloride. The undesirable sodium ions which have been linked to the resin are carried away in the fæces.

REFERENCES

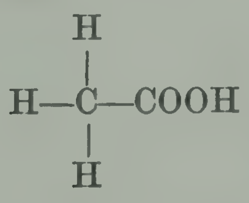
Clark, W. M. (1948). *Topics in Physical Chemistry.* Baltimore: Williams & Wilkins Co.
Mills, G. T., Munro, H. N. & Leaf, G. (1954). *Practical Biochemistry.* Glasgow: J. Smith & Son.

CHAPTER 5

THE PROTEINS

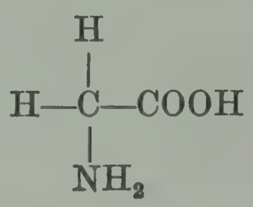
THE proteins form one of the most important as well as one of the most complex groups of biological materials comprising as they do the chief nitrogenous constituents of the tissues of the body and of the food we eat. The complete hydrolysis of various proteins yields some twenty-two different amino-acids; since there are no other products of hydrolysis it is concluded that proteins are built up exclusively of amino-acid units.

The simplest amino-acid is amino-acetic acid or glycine.



$$H{-}\overset{H}{\underset{H}{C}}{-}COOH$$

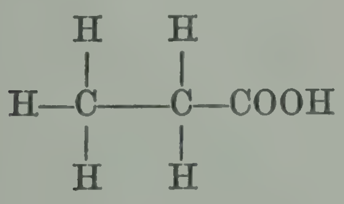
acetic acid



$$H{-}\overset{H}{\underset{NH_2}{C}}{-}COOH$$

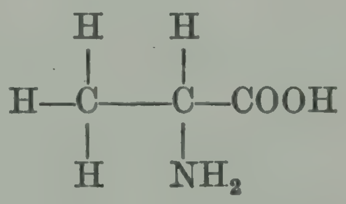
amino-acetic acid
(glycine)

The next higher amino-acid is amino-propionic acid. There are two amino derivatives of propionic acid according as the amino group is attached to the α or the β carbon atom.



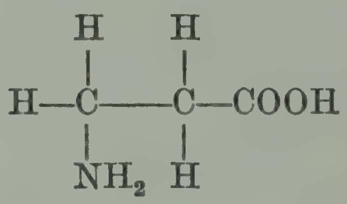
$$H{-}\overset{H}{\underset{H}{C}}{-}\overset{H}{\underset{H}{C}}{-}COOH$$

propionic acid



$$H{-}\overset{H}{\underset{H}{C}}{-}\overset{H}{\underset{NH_2}{C}}{-}COOH$$

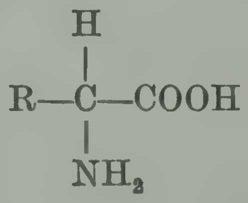
α-amino-propionic acid
(alanine)



$$H{-}\overset{H}{\underset{NH_2}{C}}{-}\overset{H}{\underset{H}{C}}{-}COOH$$

β-amino-propionic acid
(β-alanine)

In the same way butyric acid can give rise to three possible amino-acids, and so on. It so happens, however, that the amino-acids found in proteins are all α-amino-acids and they all thus have the general formula:



$$R{-}\overset{H}{\underset{NH_2}{C}}{-}COOH$$

where R is H in glycine and R is CH_3 in alanine.

It will be noted that the α-carbon atom in all amino-acids except glycine is attached to four different atoms or groups and is therefore asymmetric. The amino-acids are therefore optically active and can exist in two stereoisomeric forms. All the amino-acids which occur naturally in the tissue proteins belong to the L-series of configurationally related compounds, although some amino-acids are dextro-rotatory and others are lævo-rotatory.* Thus naturally occurring alanine is L(+) alanine, which is dextro-rotatory although it belongs to the L-series, that is it is configurationally related to L-lactic acid. Amino-acids of the D-series occur in some antibiotics, for example in gramicidin.

---

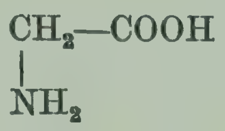
* *Cf.* p. 10.

The amino-acids obtained by hydrolysis of the proteins may be classified as follows:

Aliphatic amino-acids.

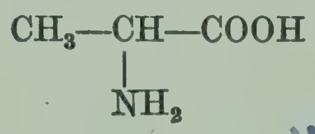
1. *Monoamino-monocarboxylic acids.*

Glycine
(amino-acetic acid)



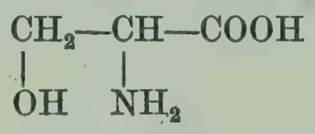
$CH_2(NH_2)-COOH$

L(+)Alanine *
(α-amino-propionic acid)



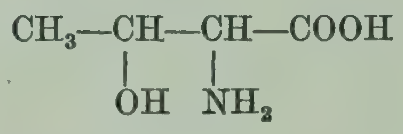
$CH_3-CH(NH_2)-COOH$

L(−)Serine
(α-amino-β-hydroxy-propionic acid)



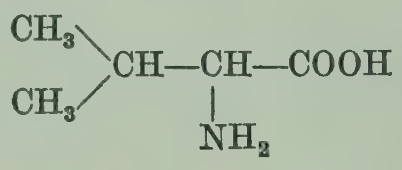
$CH_2(OH)-CH(NH_2)-COOH$

L(+)Threonine
(α-amino-β-hydroxy-butyric acid)



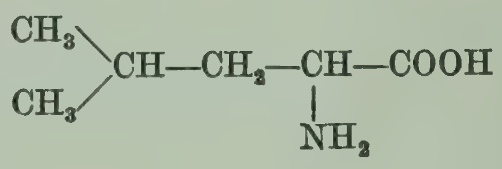
$CH_3-CH(OH)-CH(NH_2)-COOH$

L(+)Valine
(α-amino-isovaleric acid)



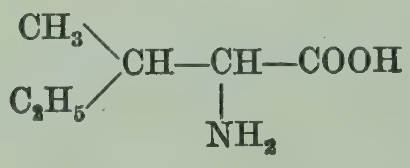
$(CH_3)_2CH-CH(NH_2)-COOH$

L(−)Leucine
(α-amino-iso-caproic acid)



$(CH_3)_2CH-CH_2-CH(NH_2)-COOH$

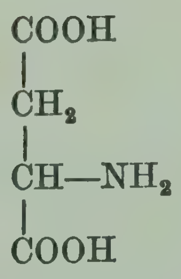
L(+)Isoleucine
(α-amino-β-methyl valeric acid)



$(CH_3)(C_2H_5)CH-CH(NH_2)-COOH$

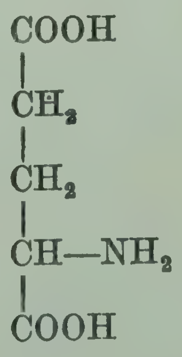
2. *Monoamino-dicarboxylic acids.*

L(−)Aspartic acid
(α-amino-succinic acid)



$COOH-CH_2-CH(NH_2)-COOH$

L(+)Glutamic acid
(α-amino-glutaric acid)



$COOH-CH_2-CH_2-CH(NH_2)-COOH$

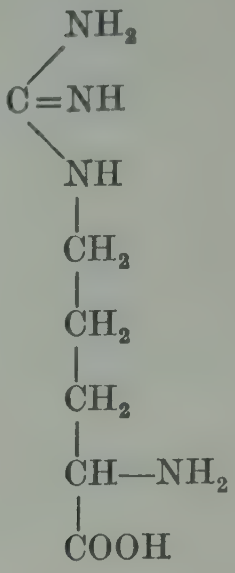
---

* The letter 'L' indicates a member of the L-series of configurationally related compounds. The + sign indicates that the acid gives a positive rotation in dilute aqueous solution.

3. *Diamino-monocarboxylic acids.*

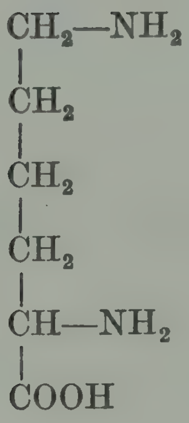
L(+)Arginine

(α-amino-δ-guanido-valeric acid)



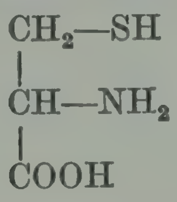
L(+)Lysine

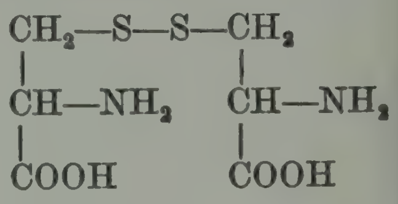
(α, ε-diamino-caproic acid)



4. *Sulphur-containing amino-acids.*

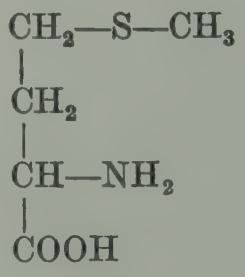
L(−)Cysteine (α-amino-β-thiol-propionic acid) and L(−)Cystine (dicysteine)





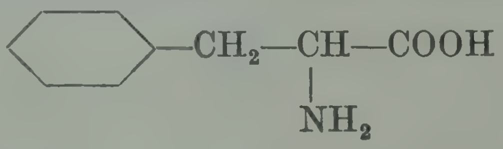
L(−)Methionine

(α-amino-γ-methylthiol-butyric acid)

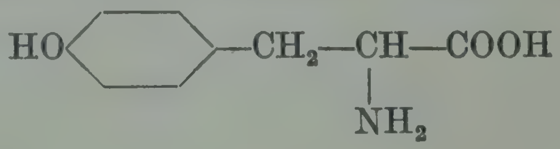


5. *Aromatic and heterocyclic amino-acids.*

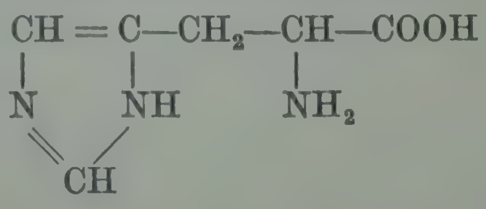
L(−)Phenylalanine (α-amino-β-phenyl-propionic acid)



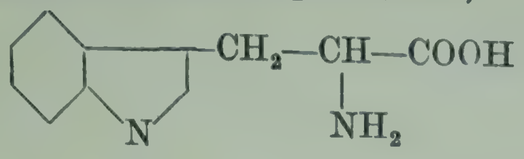
L(−)Tyrosine (α-amino-β-p-hydroxyphenyl-propionic acid)



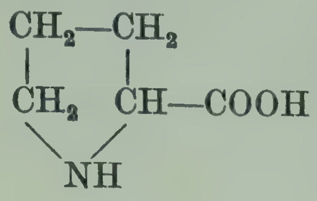
L(−)Histidine (α-amino-β-iminazolyl-propionic acid)



L(−)Tryptophan (α-amino-β-indole-propionic acid)

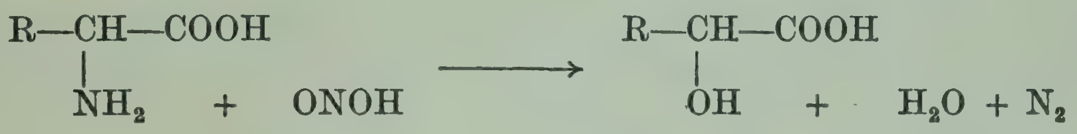


L(−)Proline (α-pyrrolidine-carboxylic acid)



Other amino-acids which are of physiological importance are ornithine and citrulline (p. 342), thyroxine and diiodotyrosine (Chapter 51), and hydroxyproline.

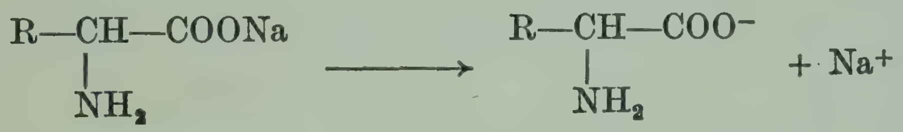
Amino-acids are colourless, crystalline solids which form esters with alcohols and salts with metals such as copper. They show certain reactions which are characteristic of the amino group; thus they react with nitrous acid with the evolution of nitrogen.



$$R-\underset{\substack{|\\ NH_2}}{CH}-COOH + ONOH \longrightarrow R-\underset{\substack{|\\ OH}}{CH}-COOH + H_2O + N_2$$

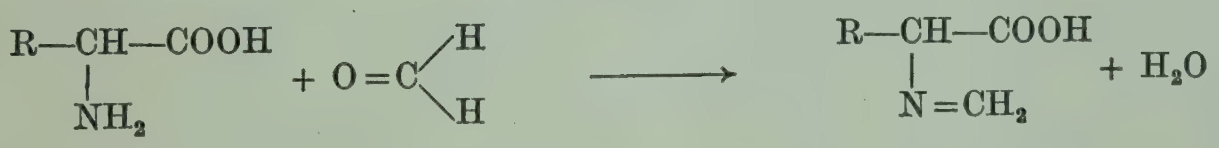
As the equation indicates, the volume of nitrogen evolved is a measure of the amount of amino-acid present.

Amino-acids like other acids react with bases, such as sodium hydroxide, to form salts which ionize thus:



$$R-\underset{\substack{|\\ NH_2}}{CH}-COONa \longrightarrow R-\underset{\substack{|\\ NH_2}}{CH}-COO^- + Na^+$$

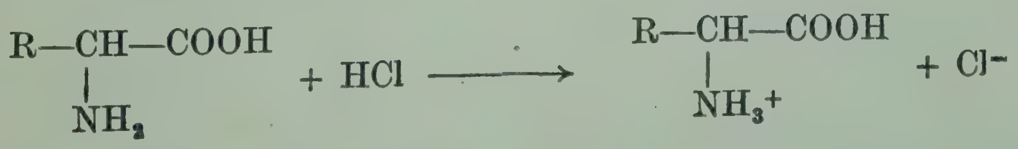
They differ however from other acids, in that they cannot be titrated directly with alkali because the basic amino group interferes with the titration. Amino-acids react with formaldehyde,



$$R-\underset{\substack{|\\ NH_2}}{CH}-COOH + O=C\begin{matrix}H\\H\end{matrix} \longrightarrow R-\underset{\substack{|\\ N=CH_2}}{CH}-COOH + H_2O$$

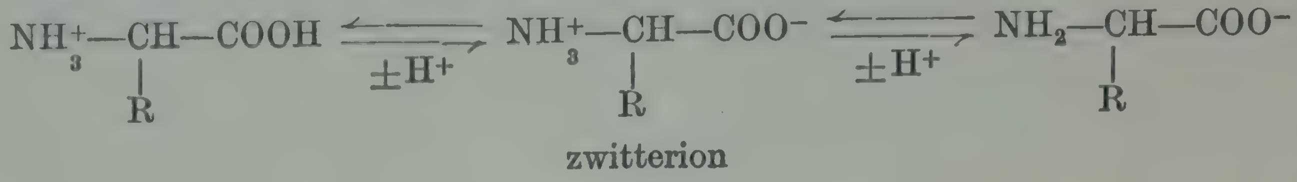
—the basic properties of the amino group being lost; the carboxyl group can then be titrated with alkali—a process known as *formol titration.*

The amino-acids by virtue of their amino groups are also bases and can therefore react with acids to form salts; for example they can react with hydrochloric acid to form hydrochlorides which ionize thus:



$$R-\underset{\substack{|\\ NH_2}}{CH}-COOH + HCl \longrightarrow R-\underset{\substack{|\\ NH_3^+}}{CH}-COOH + Cl^-$$

Since the amino-acids behave either as acids or bases they are called *amphoteric electrolytes*, or *ampholytes.* In acid solution they give rise to a positively charged ion, and in alkaline solution to a negatively charged ion. At intermediate pH values, known as the *isoelectric zone*, they yield ions carrying both positive and negative charges. Such ions are termed 'zwitterions' or 'dipolar ions' and are electrically neutral.

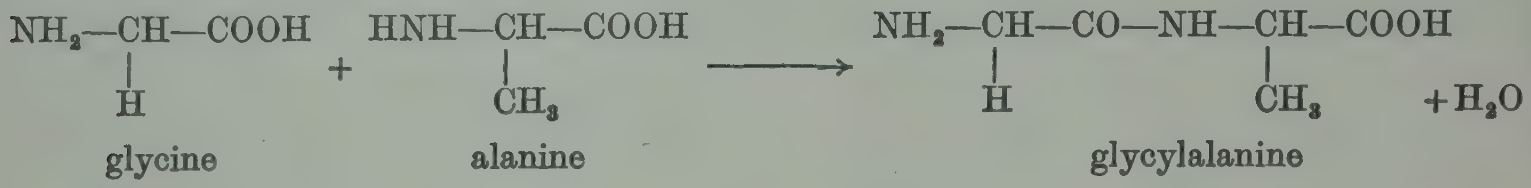


$$NH_3^+-\underset{\substack{|\\R}}{CH}-COOH \underset{\pm H^+}{\rightleftharpoons} \underset{\text{zwitterion}}{NH_3^+-\underset{\substack{|\\R}}{CH}-COO^-} \underset{\pm H^+}{\rightleftharpoons} NH_2-\underset{\substack{|\\R}}{CH}-COO^-$$

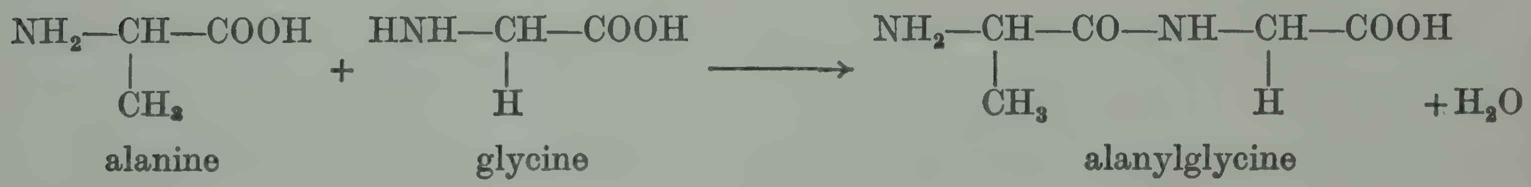
Addition of acid gives the form shown on the left and of alkali the form shown on the right.

PEPTIDES

The carboxyl group of one amino-acid can react with the amino group of another amino-acid, with the elimination of water. Thus glycine and alanine can react together in one of two ways with the formation of the glycylalanine or alanylglycine.

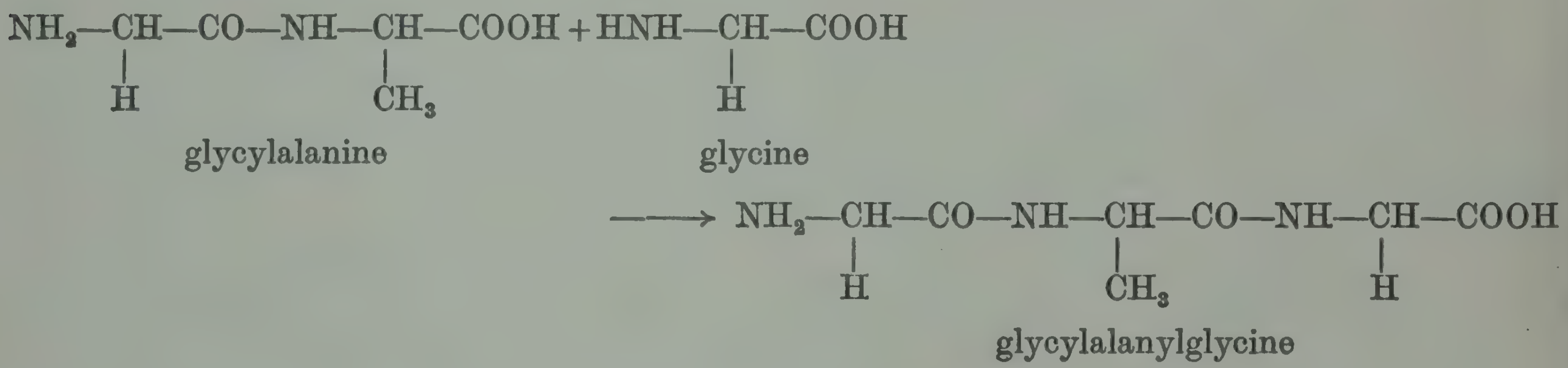


$$\underset{\text{glycine}}{NH_2-\underset{\substack{|\\H}}{CH}-COOH} + \underset{\text{alanine}}{HNH-\underset{\substack{|\\CH_3}}{CH}-COOH} \longrightarrow \underset{\text{glycylalanine}}{NH_2-\underset{\substack{|\\H}}{CH}-CO-NH-\underset{\substack{|\\CH_3}}{CH}-COOH} + H_2O$$

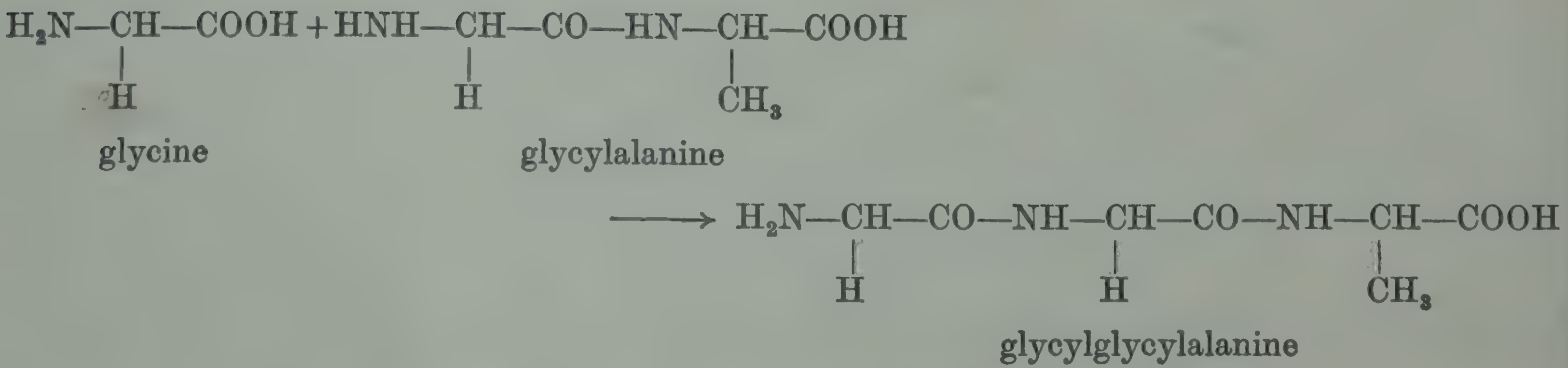


$$\underset{\text{alanine}}{NH_2-\underset{\substack{|\\CH_3}}{CH}-COOH} + \underset{\text{glycine}}{HNH-\underset{\substack{|\\H}}{CH}-COOH} \longrightarrow \underset{\text{alanylglycine}}{NH_2-\underset{\substack{|\\CH_3}}{CH}-CO-NH-\underset{\substack{|\\H}}{CH}-COOH} + H_2O$$

Glycylalanine may then react either at its amino group or at its carboxyl group with another molecule of glycine to form glycylalanylglycine or glycylglycylalanine.

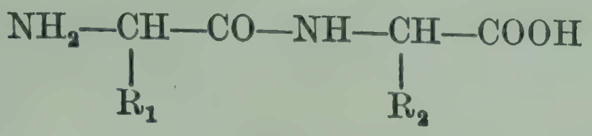


$$\underset{\text{glycylalanine}}{NH_2-\underset{\substack{|\\H}}{CH}-CO-NH-\underset{\substack{|\\CH_3}}{CH}-COOH} + \underset{\text{glycine}}{HNH-\underset{\substack{|\\H}}{CH}-COOH}$$
$$\longrightarrow \underset{\text{glycylalanylglycine}}{NH_2-\underset{\substack{|\\H}}{CH}-CO-NH-\underset{\substack{|\\CH_3}}{CH}-CO-NH-\underset{\substack{|\\H}}{CH}-COOH}$$



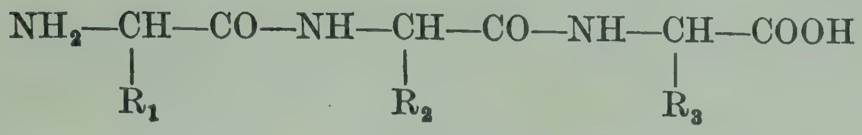
$$\underset{\text{glycine}}{H_2N-\underset{\substack{|\\H}}{CH}-COOH} + \underset{\text{glycylalanine}}{HNH-\underset{\substack{|\\H}}{CH}-CO-HN-\underset{\substack{|\\CH_3}}{CH}-COOH}$$
$$\longrightarrow \underset{\text{glycylglycylalanine}}{H_2N-\underset{\substack{|\\H}}{CH}-CO-NH-\underset{\substack{|\\H}}{CH}-CO-NH-\underset{\substack{|\\CH_3}}{CH}-COOH}$$

The compounds formed by the units of amino-acids in this way are known as *peptides* and the group —CO—NH— joining the amino-acid residues is known as the peptide linkage. When two amino-acids are joined together in this way the product is known as *dipeptide*. Three amino-acids form a *tripeptide*, and the product formed by the condensation of a large number of amino-acids is a *polypeptide*. The general formulæ for dipeptides and tripeptides are therefore as follows:



$$NH_2-\underset{\begin{array}{c}|\\R_1\end{array}}{CH}-CO-NH-\underset{\begin{array}{c}|\\R_2\end{array}}{CH}-COOH$$

dipeptide

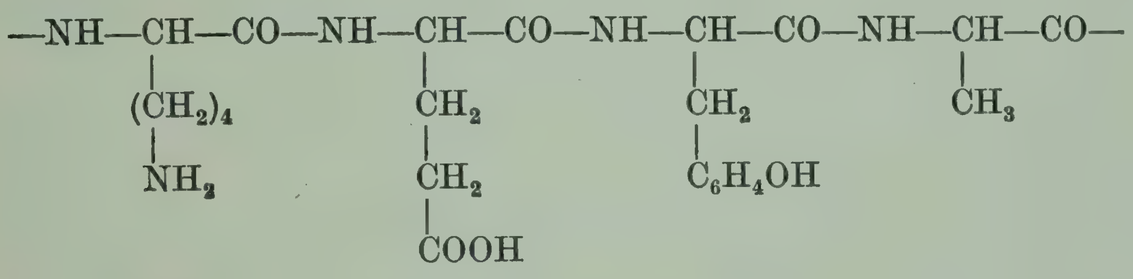


$$NH_2-\underset{\begin{array}{c}|\\R_1\end{array}}{CH}-CO-NH-\underset{\begin{array}{c}|\\R_2\end{array}}{CH}-CO-NH-\underset{\begin{array}{c}|\\R_3\end{array}}{CH}-COOH$$

tripeptide

As more than twenty amino-acids, any one of which may be used more than once, are available in nature for building up into peptides, and since the order in which they occur in the peptide chains can be varied, it will be evident that the number of possible polypeptides is enormous. For example, from 20 different amino-acids it is theoretically possible to make 400 dipeptides and 8000 tripeptides.

A section of a polypeptide chain is shown below in which the sequence of amino-acid *residues* is lysyl-glutamyl-tyrosyl-alanyl.



$$-NH-\underset{\begin{array}{c}|\\(CH_2)_4\\|\\NH_2\end{array}}{CH}-CO-NH-\underset{\begin{array}{c}|\\CH_2\\|\\CH_2\\|\\COOH\end{array}}{CH}-CO-NH-\underset{\begin{array}{c}|\\CH_2\\|\\C_6H_4OH\end{array}}{CH}-CO-NH-\underset{\begin{array}{c}|\\CH_3\end{array}}{CH}-CO-$$

Residues of lysine glutamic acid tyrosine alanine

Some peptides have important biological activity. For example glutathione, which is discussed on p. 89, is a tripeptide, glutamylcysteylglycine. Also many *antibiotics*, substances elaborated by some micro-organisms which inhibit the growth of other micro-organisms, are peptide in nature. Some of them, indeed, contain residues of rare amino-acids or amino-acids of the D-series.

PROTEIN STRUCTURE

The protein molecule may be regarded as a very elaborate polypeptide consisting of a large number of amino-acid residues. In discussing the structure of any protein it is essential to have information about (*a*) its molecular size, (*b*) the nature and relative proportions of the various constituent amino-acids, (*c*) the sequence of these acids in the chain, and (*d*) the way in which the chain is ultimately arranged in three dimensions.

The Molecular Size of Proteins.—The molecular weight or, more correctly, the particle size of most proteins is very high (Table 5, 1) and cannot be determined by the usual methods of organic chemistry such as depression of the freezing point. The most effective methods at present in use depend upon accurate measurements either of osmotic pressure or of sedimentation velocity and sedimentation equilibrium in the ultracentrifuge. The relative particle sizes of some blood proteins are shown in Fig. 23, 9.

The Nature and Proportions of the Constituent Amino-Acids.—By modern methods of paper chromatography it is possible to identify the amino-acids present in a hydrolysate of as little as 1 mg. of protein. With larger amounts of protein (10 to 20 mg.) it is possible to subject a protein hydrolysate to chromatography in columns of starch or of ion-exchange resins (Ch. 4) in order to separate the individual amino-acids and to determine the relative proportions of each by colorimetric methods.

TABLE 5, 1.—*Molecular Weights of Proteins*

Protein	Molecular Weight
Ribonuclease	12,700
Lactalbumin	17,400
Gliadin	27,500
Lactoglobulin	41,500
Egg albumin	45,200
Hæmoglobin (horse)	68,000
Serum albumin (horse)	70,000
Antipneumococcus serum globulin (rabbit)	157,000
Serum globulin (horse)	167,000
Antipneumococcus serum globulin (man)	195,000
Catalase	250,000
Urease	480,000
Thyreoglobulin (pig)	630,000
Antipneumococcus serum globulin (horse)	920,000

Some proteins contain all the common amino-acids, but in others certain amino-acids may be missing while other amino-acids may be present in relatively large amounts. The amino-acid content of some common proteins is shown in Table, 5, 2. Gelatin, for example, contains neither tyrosine nor tryptophan, and zein no lysine or tryptophan. Glycine is however abundant in gelatin, cysteine in keratin, and proline in collagen and gelatin.

The Sequence of Amino-Acids.—It should be clear that at one end of the long polypeptide chain forming the protein molecule there is an amino-acid residue with a free amino group and at the other end an amino-acid residue with a free carboxyl group. These are called the *end-groups*.

Moreover, whenever diamino or dicarboxylic acids occur in the polypeptide chain, elaboration of the chain by branching is theoretically possible and each branch has, of course, an end-group.

The nature of the end-groups may be ascertained by suitably labelling them and then finding after hydrolysis which amino-acid possesses the label.

TABLE 5, 2

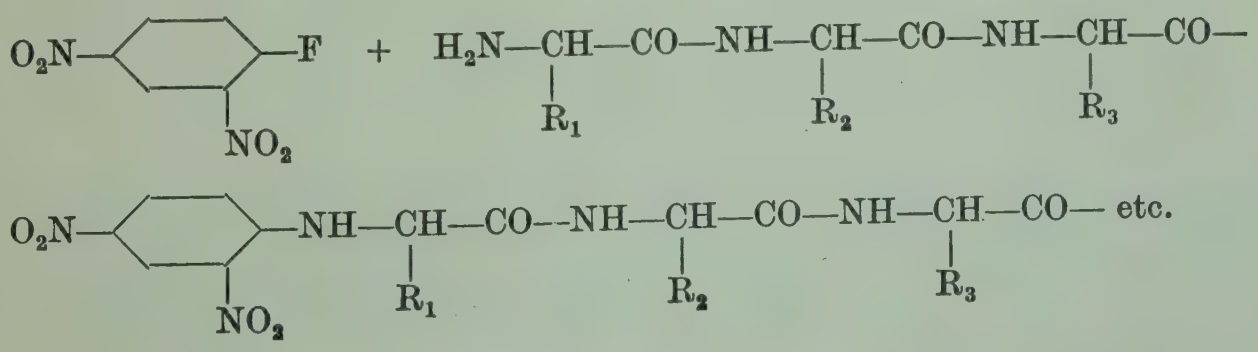
The Amino-Acid Composition of Some Proteins (g./100 g. protein)

	Serum Albumin	Insulin	Lacto-globulin	Gelatin
Glycine	5·6	4·3	1·5	26·9
Alanine	3·7	4·5	6·42	9·3
Valine *	4·1	7·75	5·72	3·3
Leucine *	7·1	13·2	15·46	5·23
Isoleucine *	4·8	2·77	5·90	
Proline	5·7	2·53	5·4	14·8
Phenylalanine *	4·6	8·14	3·95	2·55
Cystine	2·3	12·5	2·29	..
Cysteine	0·4	..	1·10	
Arginine *	7·8	3·07	2·86	8·55
Histidine *	2·6	4·91	1·58	0·73
Lysine *	9·2	2·51	11·3	4·6
Aspartic acid	13·1	6·8	11·4	5·6
Glutamic acid	14·5	18·6	21·5	11·2
Serine	7·0	5·23	4·07	3·18
Threonine *	6·1	2·08	5·15	2·20
Tyrosine	5·5	13·03	3·73	1·0
Tryptophan *	3·3	..	1·90	..
Methionine *	2·6	..	2·20	0·9
Hydroxyproline	..	..	..	14·5

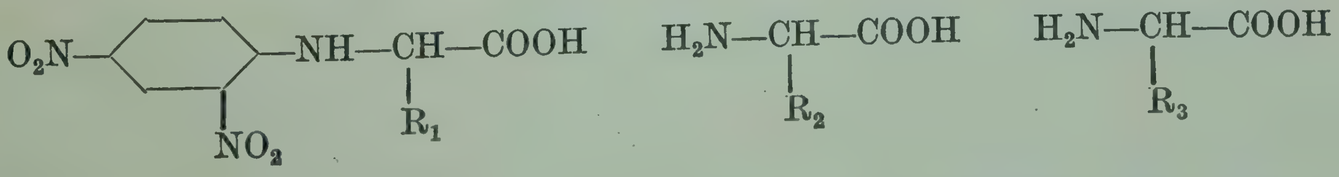
* Nutritionally essential amino-acids.

From G. T. Mills, H. N. Munro & G. Leaf (1954). *Practical Biochemistry*, based on data from G. R. Tristram (1949). *Advances in Protein Chemistry*, **5**, 83.

For example, a free amino group at the end of a chain may be labelled with dinitrofluorobenzene thus:—


O_2N
F
+
$H_2N—CH—CO—NH—CH—CO—NH—CH—CO—$
R_1
R_2
R_3
NO_2
O_2N
$—NH—CH—CO—NH—CH—CO—NH—CH—CO—$ etc.
R_1
R_2
R_3
NO_2

When hydrolysis is subsequently carried out and the products separated



the terminal amino-acid is isolated as the dinitrophenyl derivative.

In such ways as this it is possible to determine the nature of the end-groups and their number, and hence to obtain an estimate of the number of side chains per molecule.

Moreover by partial hydrolysis of the protein to di- and tripeptides which can be separated chromatographically it is possible, in some cases, to work out the order in which the amino-acids occur in the polypeptide chain.

The Spatial Arrangement of the Chains.—The work of Astbury and his colleagues in Leeds, using the method of X-ray analysis, has revealed that from the structural point of view proteins may be divided into two main groups: (1) *the fibrous proteins* and (2) *the corpuscular* or *globular proteins.*

The fibrous proteins are used by the body chiefly for structural purposes. Most of them are extracellular and insoluble in water.

There are two kinds of *fibrous proteins*:

(*a*) The main feature of proteins in the *keratin-myosin-fibrinogen group* is their elasticity. The structural unit is a long polypeptide chain which is in equilibrium in a characteristic folded or coiled form. Side chains, characteristic of the component amino-acids, project at right angles so that each molecule possesses a characteristic pattern or design according to the nature of these particular amino-acids. Keratin, for example, normally exists in the α form but when the chains are stretched an extended zigzag pattern is obtained in which keratin is said to be in the β state (Fig. 5, 3). When the tension is released the molecule returns to the original state and α keratin is re-formed. If the keratin of hair is heated, for example by steam, while it is in the stretched state, it does not go

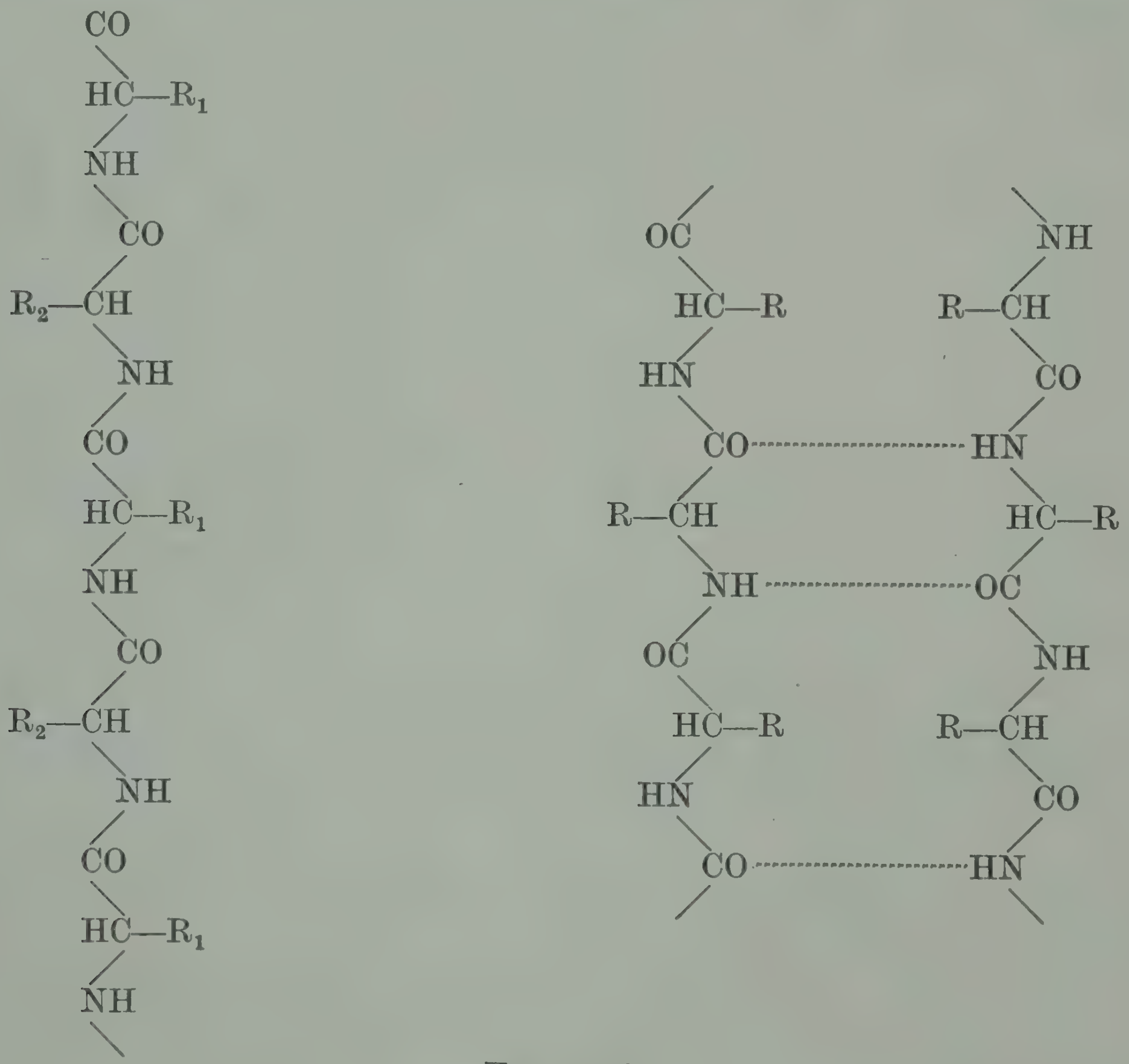
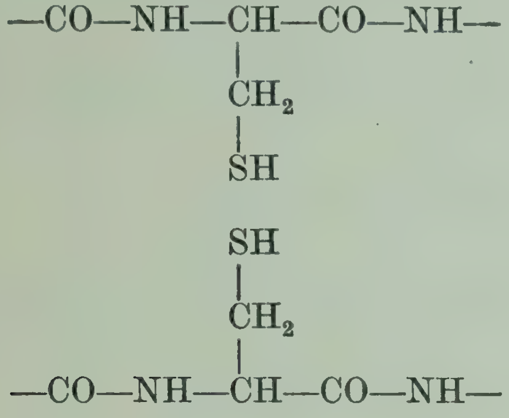


FIG. 5, 3

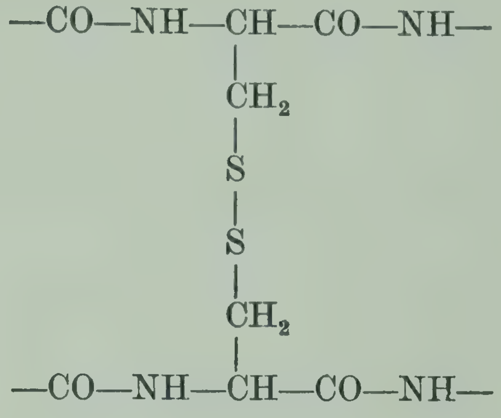
(Left) Astbury's suggested arrangement of the β form of the polypeptide chain in fibrous proteins. The C, N and O atoms lie in one plane with the R_1 and R_2 groupings projecting above and below it respectively. (Right) The linking of adjacent chains by hydrogen bonds.

back to the α form when the tension is reduced but remains 'set' in the β form. This fact is made use of in the so-called 'permanent waving' of hair.

It should be emphasized that the molecular shape of protein molecules depends to a large extent on weak cross-linkages between active groups in the side chains of the amino-acids in the main peptide chains. For example the side chains of adjacent cysteine residues may become linked by removal of hydrogen, thus:



$$\begin{array}{c} \text{—CO—NH—CH—CO—NH—} \\ | \\ CH_2 \\ | \\ SH \\ \\ SH \\ | \\ CH_2 \\ | \\ \text{—CO—NH—CH—CO—NH—} \end{array}$$



$$\begin{array}{c} \text{—CO—NH—CH—CO—NH—} \\ | \\ CH_2 \\ | \\ S \\ | \\ S \\ | \\ CH_2 \\ | \\ \text{—CO—NH—CH—CO—NH—} \end{array}$$

Other types of linkage are also found, the most important of which is the 'hydrogen bond' illustrated in Fig. 5, 3 (right) in which the hydrogen atom attached to the nitrogen in one peptide bond is attracted to the oxygen atom in an adjacent peptide bond.

According to Pauling the polypeptide chain in such proteins as α-keratin is coiled in the form of a spiral or helix with the main axis parallel to the fibre axis and with 3·7 amino-acid units per turn, the shape of the helix being maintained by hydrogen bonds. In β-keratin the parallel peptide chains lie side by side forming what Pauling calls the 'pleated sheet structure' with adjacent chains linked by hydrogen bonds.

(*b*) Proteins of the *collagen group* include the white fibres of connective tissue, tendons and cartilage, and the scales and fins of fishes. These proteins form fibres which do not normally show long-range elasticity until they have been subjected to the action of heat or swelling agents.

The *globular* or *corpuscular proteins* include many of the common water soluble proteins such as those of blood serum and egg albumin. Much less is known about them than about the fibrous proteins, but they probably consist also of polypeptide chains folded in characteristic patterns, forming what Astbury has termed 'large rounded and usually unstable packets that have so far given no direct clue at all as to how they are shaped inside.' The egg albumin molecule, for example, is believed to consist of four superimposed laminæ of folded polypeptide chains. Under suitable conditions it is possible for globular proteins to be converted to the fibrous type.

The protein molecule about which we have the most complete knowledge is insulin which has been examined in detail by Sanger and his colleagues. The insulin molecule consists of two kinds of polypeptide chain, the A type and the B type, linked together by disulphide bridges as described above. The amino-acids present are shown in Table 5, 2, and the sequence of amino-acids in the A chain is Gly. Ileu. Val. Glu. Glu. Cys. Cys. Ala. Ser. Val. Cys. Ser. Leu. Tyr. Glu. Leu. Glu. Asp. Tyr. Gys. Asp., and in the B chain is Phe. Val. Asp. Glu. His. Leu. Cys. Gly. Ser. His. Leu. Val. Glu. Ala. Leu. Tyr. Leu. Val. Cys. Gly. Glu. Arg. Gly. Phe. Phe. Tyr. Thr. Pro. Lys. Ala. (where these contractions stand for the appropriate amino-acids, i.e. Gly. = a glycine residue).

The two types of chain are joined together thus:

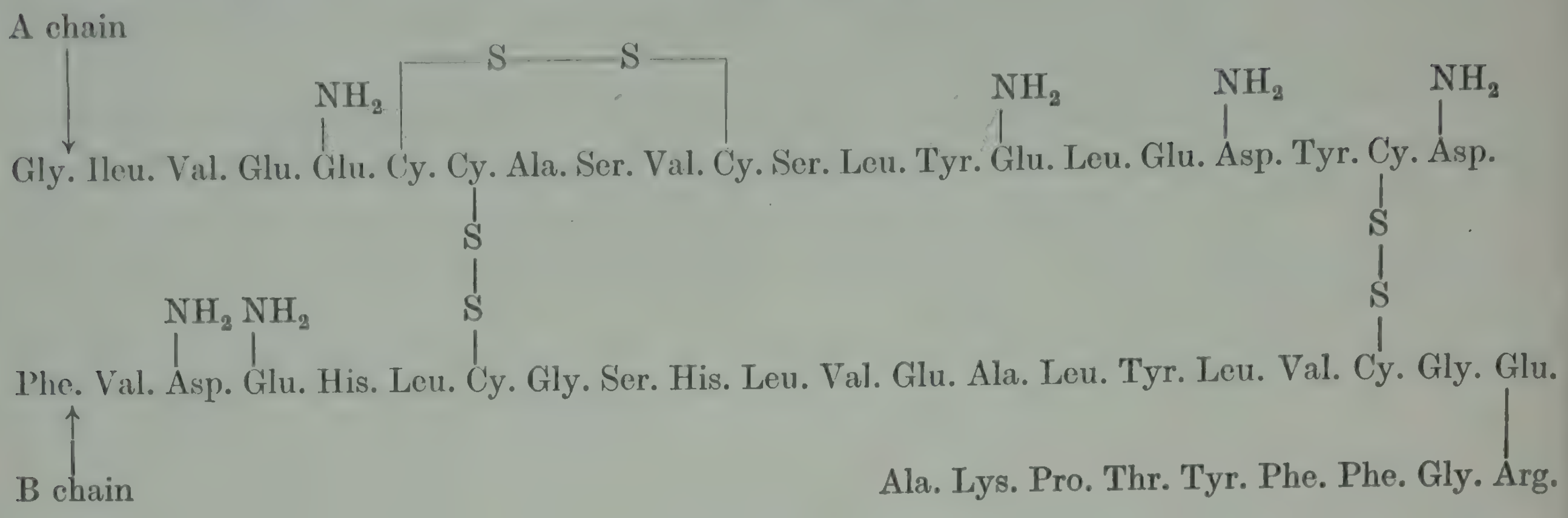

A chain
Gly. Ileu. Val. Glu. Glu. Cy. Cy. Ala. Ser. Val. Cy. Ser. Leu. Tyr. Glu. Leu. Glu. Asp. Tyr. Cy. Asp.
S
S
NH₂
Phe. Val. Asp. Glu. His. Leu. Cy. Gly. Ser. His. Leu. Val. Glu. Ala. Leu. Tyr. Leu. Val. Cy. Gly. Glu.
Ala. Lys. Pro. Thr. Tyr. Phe. Phe. Gly. Arg.
B chain

FIG. 5, 4

Sequences of amino-acids in the insulin molecule (Sanger).

the cysteine which occurs as 7th and 20th residues in the A chain linking with the cysteine which occurs as 7th and 19th residues in the B chain. An additional disulphide bridge is formed between the cysteine residues in positions 6 and 11 in the A chain. The amino-acid residues with the free amino group are shown on the left.

Colloidal Properties of Proteins.—Many proteins have been purified to a high degree and even obtained in crystalline form. This, however, can no longer be accepted as a criterion of purity of a protein, since in some cases the crystalline material has now been shown to be a mixture. Although proteins can in many cases be crystallized, their solutions are typical examples of the lyophilic or emulsoid type of colloid in which the behaviour of the solution depends to a great extent upon its pH. At one end of the chain of amino-acids forming the protein molecule there is a free amino group and at the other end a free carboxyl group, while other free amino and carboxyl groups may occur at intervals along the chain at the sites occupied by diamino or dicarboxylic acids. Proteins, like the amino-acids, are therefore amphoteric substances and can exist in the form of 'zwitter-ions' (p. 44). In acid solution the protein ions are positively charged, while in alkaline solution they are negatively charged. At some intermediate pH value, known as the *isoelectric point*, the ions are electrically neutral. At this pH, which is characteristic for each protein, the solution is least stable. The protein is therefore most readily removed from solution at its isoelectric point, either by precipitation or by adsorption. At the isoelectric point the viscosity of the protein solution is minimal.

By virtue of the presence of an electric charge on the particles, proteins migrate in an electric field. This property is made use of in the important technique of *electrophoresis* evolved by Tiselius. The apparatus consists of a glass U-tube in which the protein solution is placed so as to fill each limb to the half-way mark (Fig. 5, 5). A buffer solution at pH 8·6 is then added to fill each limb to the top and electrodes connected to a battery are inserted. The charged protein particles move towards the positive pole under the influence of the electric field, but the different types of protein move at different rates, so that a succession of boundaries is formed moving towards the positive pole. These boundaries, each corresponding to a different protein, can be photographed by special apparatus which records the electrophoretic pattern of the blood plasma or protein solution. The pattern given by normal blood plasma is shown in Fig. 5, 6. Since

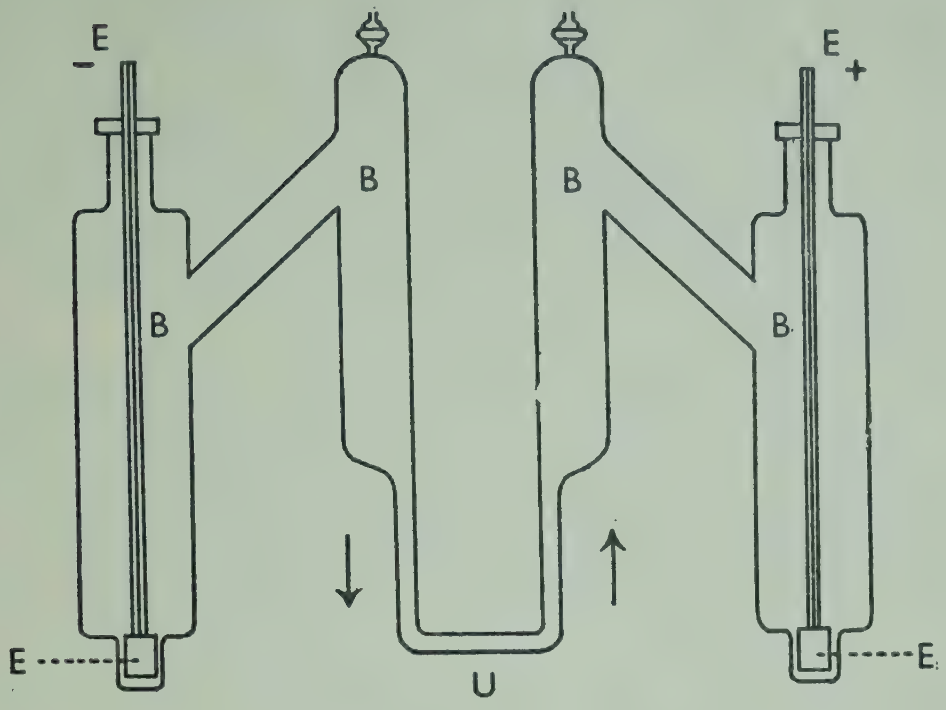


FIG. 5, 5

Diagrammatic representation of Tiselius electrophoresis cell. The optical mechanism for recording the moving boundaries is not shown. The arrows indicate the direction of movement of the plasma proteins at physiological pH values.

U = U-tube containing protein solution.

B = chambers containing buffer solution.

E = non-polarizable electrodes.

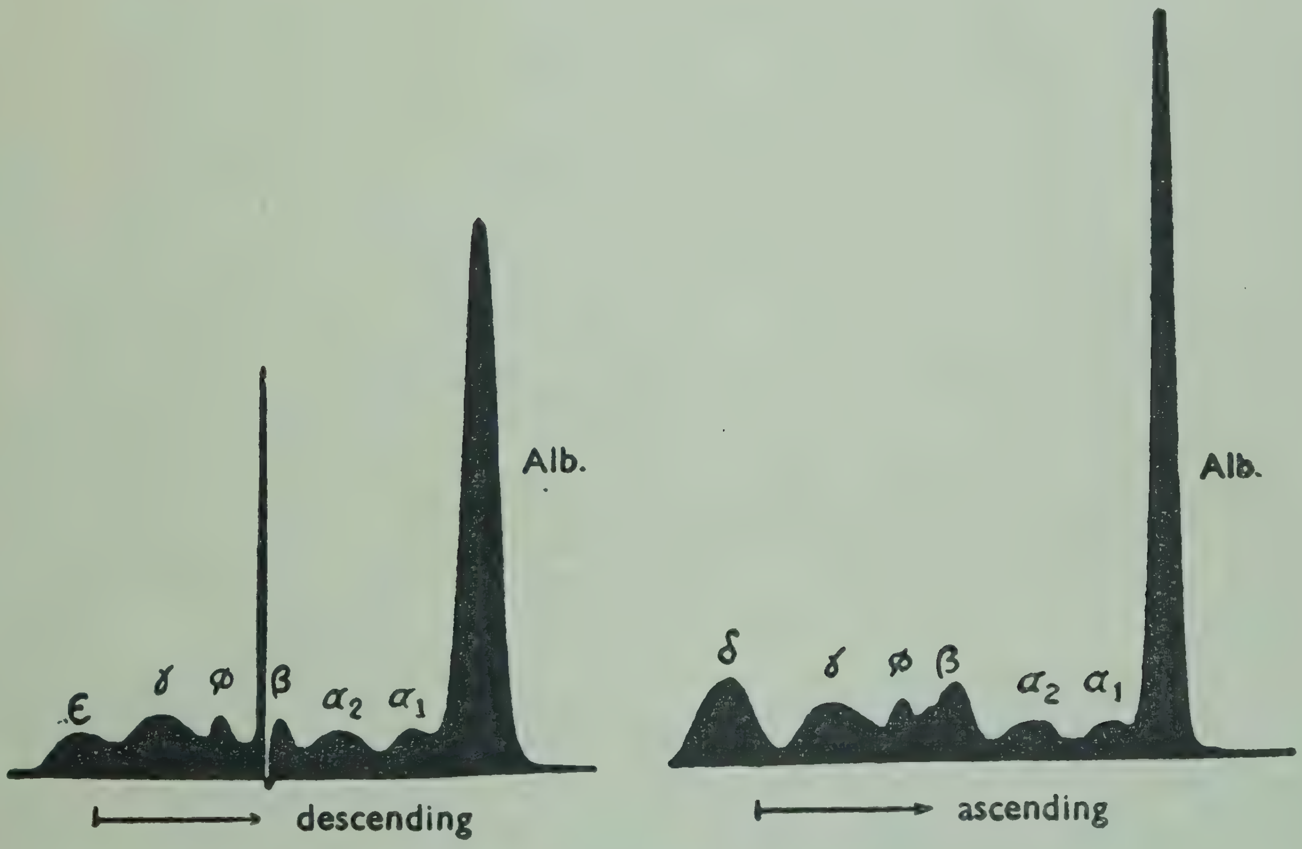


FIG. 5, 6

Electrophoretic pattern of normal human plasma. The boundaries are moving downwards in the descending limb and upwards in the ascending limb. Each peak on the diagram represents a different moving boundary, that is a separate protein component. The peak for albumin is most marked and is moving most rapidly (V. P. Dole (1944). *J. clin. Invest.*, **23,** 708).

FOOD TECHNOLOGICAL RESEARCH
30 JUN 1958
LIBRARY

electrophoresis can reveal fine details of the protein constituents this apparatus has proved to be of value for studying alterations in the plasma protein occurring in disease. A large-scale version of this apparatus can be employed to prepare the different protein fractions in a relatively pure state by making use of their different mobilities in the electric field.

The Tiselius electrophoresis apparatus is expensive and complicated but a recent development, which enables the process to be carried out much more simply, is electrophoresis on paper.

In this technique a small volume (0·05 ml.) of a protein solution is applied as a narrow transverse band at the centre of a strip of filter paper about 50 cm long and 3 cm broad. The paper is moistened with suitable buffer and is held between glass plates with the projecting ends dipping into reservoirs of buffer in which are placed electrodes attached to a suitable source of direct current. The proteins migrate under the influence of the electrical field and after a suitable time (e.g. 18 hours) the paper is removed, dried and treated with an alcoholic solution of mercuric chloride containing a small amount of bromophenol blue. The paper is then washed to remove excess reagent when the protein-containing areas show up as blue bands against a white background. This technique is now employed for the examination of serum in clinical work and the type of pattern obtained is illustrated in Fig. 5, 7. The intensity of the colour bands is approximately proportional to the amount of protein in each fraction, and when suitable precautions are taken the method can be made roughly quantitative.

Denaturation.—When alcohol is added to a solution of a soluble protein the protein is precipitated. If the precipitate is immediately filtered off and treated with water it redissolves but if it is left in contact with the alcohol for a short period of time it will no longer redissolve when treated with water. The protein has obviously undergone some change which alters its solubility and it is now described as being *denatured*. Proteins in aqueous solution may be denatured by a number of physical or chemical agents, such as heat, ultraviolet light, acid, alkali, or alcohol. Denatured proteins are in general less soluble than the naturally occurring or *native* proteins from which they are derived. They tend to precipitate out much more readily at their isoelectric points, although the denatured proteins may be soluble in strongly acid or strongly alkaline solutions. Thus, if a solution of albumin is denatured by heat at or near the isoelectric point the denatured protein precipitates out. On the other hand, if albumin is denatured by heat in strongly alkaline solution the denatured protein remains in solution and does not precipitate out until the solution has been acidified to a value near that of the isoelectric point. Such a precipitate of denatured protein is readily soluble in strong alkali and although the coagulation or precipitation of denatured proteins at the isoelectric point is a reversible process, denaturation itself is irreversible and most proteins once denatured cannot be restored to the native state. The exact changes involved in denaturation are complex but probably involve rupture of the weak cross-linkages between the active groups in the side chains of the amino-acid residues in the main polypeptide chains, so that the characteristic pattern of the molecule is lost. Denatured proteins are devoid of many of the characteristic properties of the individual native proteins from which they have been derived. Thus denatured proteins cannot be crystallized and they do not possess specific immunological properties. If a protein with enzyme activity is denatured the characteristic enzymic activity is lost.

The Classification of the Proteins.—Proteins are usually classified as (*a*) the simple proteins and (*b*) the conjugated proteins. The simple proteins on hydrolysis yield only amino-acids and they include:

Albumins
Globulins
Glutelins
Gliadins
Scleroproteins
Protamines
Histones

A conjugated protein is a protein to which is attached some other substance known as the prosthetic group. The conjugated proteins include:

Phosphoproteins
Glycoproteins
Nucleoproteins
Chromoproteins
Lipoproteins

This classification is quite arbitrary but is convenient for descriptive purposes.

THE SIMPLE PROTEINS.—(1) *The Albumins.*—The albumins are simple proteins which contain nearly all of the known amino-acids. They are found, for example, in blood plasma and in egg-white. They are soluble in water and in dilute solutions of salts and can be coagulated by heat; in other words when albumin solutions are heated near the isoelectric point the protein is denatured and the denatured protein precipitates out in the form of a coagulum. This fact is made use of in testing for the presence of albumin in urine. Albumins are precipitated from solution by full saturation with ammonium sulphate. In the natural state they are frequently associated with small amounts of mucopolysaccharides (p. 17).

(2) *The Globulins.*—The globulins are found in nature associated with the albumins, for example in the blood plasma. They are also heat-coagulable and soluble in dilute solutions of salts, but differ from the albumins by being insoluble in water. They can be precipitated from solution by half-saturation with ammonium sulphate or by full saturation with magnesium sulphate.

(3) *The Glutelins and Gliadins.*—The glutelins and gliadins are plant proteins which are found together in cereals. They are insoluble in water but when kneaded with water they form a tenacious sticky mass known as dough.

(4) *The Scleroproteins* are insoluble in water or in salt solutions. They form the principal organic constituent of the connective tissue and the external coverings of animals. They include the *keratin* of hair, horns, feathers, the *collagen* of cartilage and of the white fibrous connective tissue, and the *elastin* of the yellow elastic fibres. When collagen is boiled for a long time with water it is converted into the protein *gelatin*.

(5) *The Protamines.*—The protamines are very simple proteins of relatively low molecular weights (2000 to 4000) found in the ripe sperm cells of certain fish. They contain only a few amino-acids but because they are exceptionally rich in the amino-acid arginine the protamines are strongly basic in character. By virtue of their basic properties the protamines are able to combine with other substances, for example the nucleic acids, and

they may even be made to combine with certain other proteins such as insulin. Protamine insulin, which is less soluble than ordinary insulin, is extensively used in the treatment of diabetes mellitus.

(6) *The Histones.*—The histones resemble the protamines in being of low molecular weight but their amino-acid content is more varied than that of the protamines, and unlike the protamines they contain aromatic amino-acids. The histones are found associated with nucleic acid in the nucleoproteins of the cell nuclei and can be conveniently obtained from spleen, thymus, or the nucleated red blood cells of birds.

CONJUGATED PROTEINS.—In the *phosphoproteins* the protein molecule is bound to phosphoric acid which can however be easily split off as inorganic phosphate by treatment with dilute sodium hydroxide. The chief phosphoproteins are caseinogen of milk and vitellin of egg-yolk.

The *glycoproteins* contain carbohydrate bound to a protein molecule. The carbohydrate is usually one of the complex mucopolysaccharides. They have been discussed on p. 18.

The *nucleoproteins* are compounds of nucleic acid with protein. They are discussed in Chapter 6.

The *chromoproteins* are compounds of protein with a coloured material. Hæmoglobin, for example, contains the protein globin and the red iron-containing pigment hæm. Hæmocyanins, which are respiratory pigments found in the blood of some invertebrates, are chromoproteins containing copper.

Lipoproteins are compounds of protein with lipids. The lipid cannot be extracted with ether so long as the protein remains in the native state. When the protein has been denatured, for example by alcohol, the lipid portion can easily be dissolved out with ether. A considerable proportion of the lipids of the blood plasma appear to exist in the form of lipoproteins (see p. 371).

The General Reactions of Proteins.—(1) HEAT COAGULATION.—Proteins of the albumin and globulin class are readily denatured by heat and tend to precipitate out when their solutions are boiled, since the denatured protein is insoluble at its isoelectric point. In carrying out this test it is customary to make certain that the solution is slightly acid before boiling. If the protein is heated in alkaline solution the denatured protein remains in solution.

(2) PRECIPITATION REACTIONS.—Soluble proteins can in general be precipitated from solution by the addition of (*a*) neutral salts such as ammonium sulphate (half saturation to precipitate globulins and full saturation to precipitate albumins), (*b*) alcohol, (*c*) strong mineral acids, e.g. nitric acid, (*d*) salts of heavy metals, (*e*) the so-called alkaloidal reagents, e.g. phosphotungstic acid, sulphosalicylic acid, picric acid and tannic acid. The use of these precipitants is described in textbooks of practical biochemistry.

(3) COLOUR REACTIONS.—There are many colour reactions for proteins, but most of them are tests for specific groupings, or for certain amino-acids, rather than for proteins themselves. They are therefore given only by proteins containing such groupings or amino-acids.

(a) *The Biuret Test.*—The biuret test is given by all substances which contain at least two peptide (—CO—NH—) groupings. It is therefore given by proteins and by polypeptides, as well as by the simple substance biuret itself:

$$\underset{\text{biuret}}{NH_2\text{—}CO\text{—}NH\text{—}CO\text{—}NH_2}$$

On the addition of strong alkali and a small amount of very dilute copper sulphate to the protein solution, a characteristic violet colour is obtained.

(b) *The Xanthoproteic Reaction* is given by all proteins containing amino-acids with a benzene ring. *Millon's reaction* is a test for tyrosine or proteins containing tyrosine, and the *Glyoxylic reaction* is a test for tryptophan or proteins containing it.

'The Digestive Derivatives' of Proteins.—When proteins are completely hydrolysed by treatment with acid, or by means of enzymes, the ultimate products are the amino-acids. But a series of intermediate compounds between the proteins and the amino-acids has also been described, and in the older literature sharp distinctions were drawn between these compounds, which were known as *metaproteins*, *proteoses*, *peptones* and *polypeptides*. It is very doubtful if these distinctions should be maintained, since it appears that most of these so-called digestive derivatives are complex mixtures which cannot easily be classified into definite categories. The higher members of the series, such as the metaproteins, are more easily precipitated from solution by such materials as ammonium sulphate than are the peptones or polypeptides, but the distinction between the various members of the group is not sharp.

The Biological Reactions of the Proteins.—A method capable of detecting differences between certain proteins which are very similar in chemical and physical properties makes use of biological rather than chemical reactions. Although this matter is very complex, and cannot be treated here in detail, it depends essentially upon the fact that most animal proteins are specific to the species which produce them and behave as 'foreign' substances to other species. Thus, although the serum globulins from different species of mammals cannot be distinguished from each other by either chemical or physical means, differences between them may be detected by what is known as a *precipitin* test. If a rabbit, for example, is given repeated small injections of some 'foreign' protein (the antigen) such as globulin prepared from human serum, a specific substance or antibody, probably another protein, is produced in the animal's body and circulates in its blood. If some of the blood serum of a rabbit sensitized in this way is then added to a solution of human serum proteins a precipitate is formed, but no reaction will be obtained with a solution of the serum proteins of any other species. In this way it is possible to identify the species from which a sample of blood has been obtained and a test of this sort is of great value in medico-legal work. A similar process occurs when the body is invaded by bacteria. The normal reaction of the tissues is to produce protein antibodies which are specific for the proteins or antigens of the invading organism. Thus the blood serum of a patient suffering from pneumonia contains antibodies specific for the *Pneumococcus*.

Further discussion of these *immunological* properties of the proteins is beyond the scope of this book, but the subject is mentioned again briefly on p. 378 in relation to the problem of blood groups.

Proteins of different species of animals may also be distinguished by eliciting what are termed *anaphylactic reactions*. Thus if a guinea-pig is sensitized by the injection of a small amount of foreign protein, such as albumin from hens' eggs, a second injection of the same foreign protein given some days later may cause the development of a severe reaction which may culminate in death. The phenomenon is known as anaphylactic shock. This reaction also possesses a high order of specificity; if, for example, the second injection contains protein from ducks' eggs no anaphylactic shock occurs.

REFERENCES

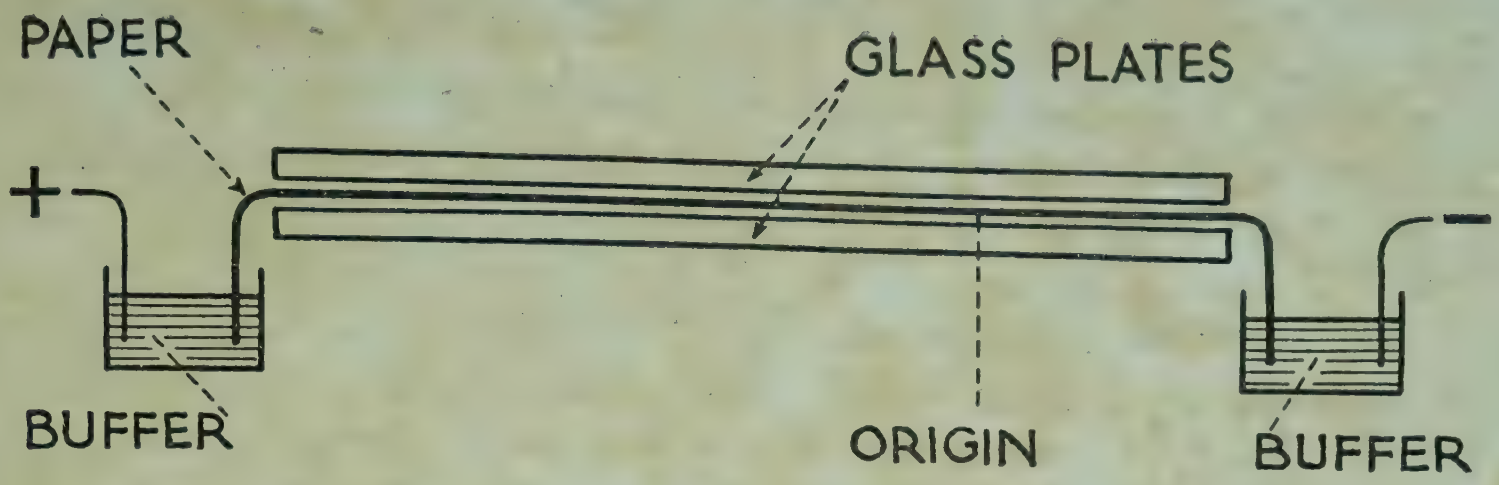
Greenberg, D. M. (1951). *Amino-Acids and Proteins.* Springfield, Ill.: C. C. Thomas.

Haurowitz, F. (1950). *Chemistry and Biology of Proteins.* New York: Academic Press.

Mills, G. T., Munro, H. N. & Leaf, G. (1954). *Practical Biochemistry*, 2nd ed. Glasgow: John Smith.

Neurath, H. & Bailey, K. (1953). *The Proteins.* New York: Academic Press.

Tracy, M. V. (1948). *Proteins and Life.* London: The Pilot Press.


PAPER
GLASS PLATES
+
−
BUFFER
ORIGIN
BUFFER

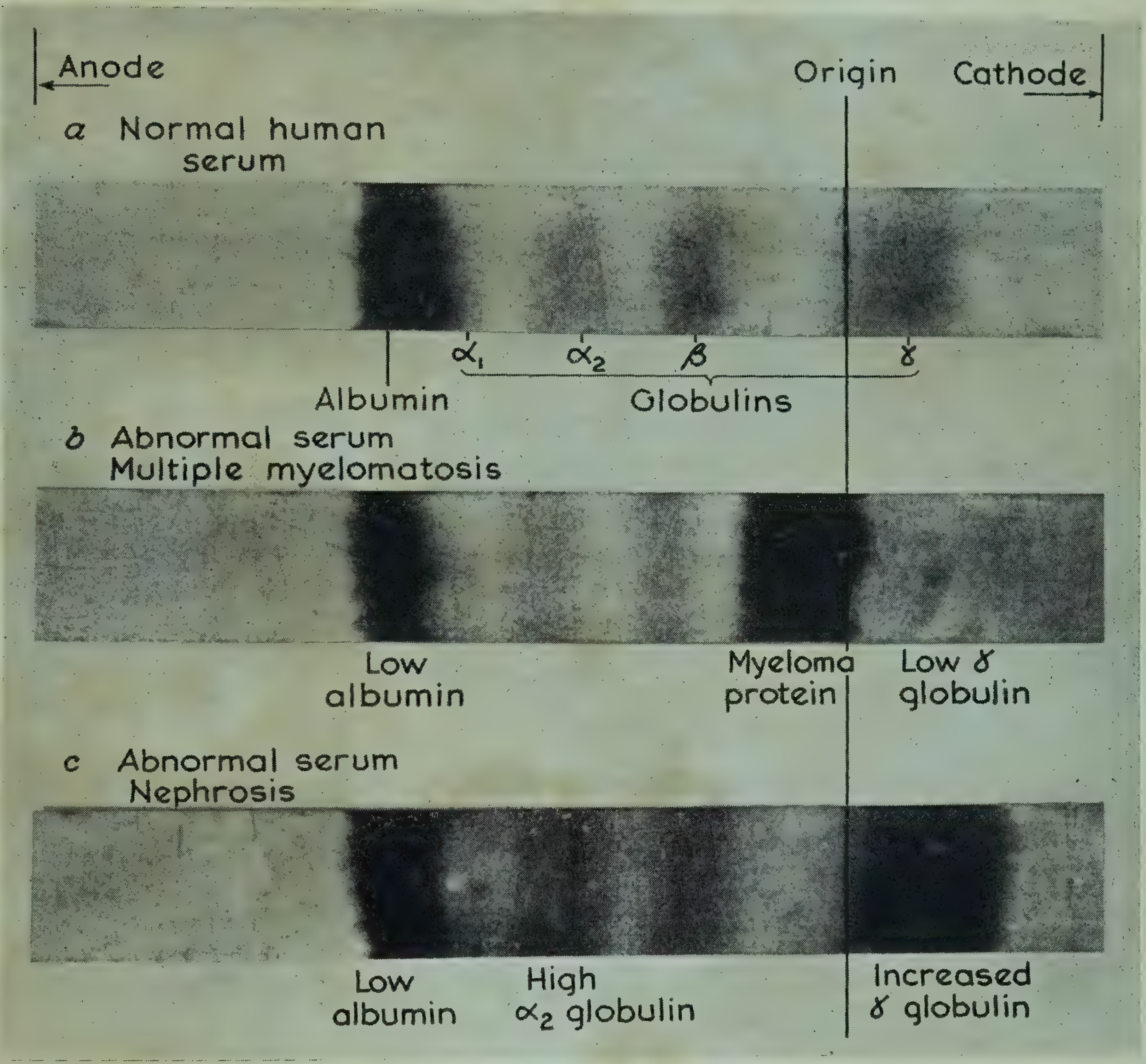

Anode
Origin
Cathode
a Normal human serum
α_1
α_2
β
γ
Albumin
Globulins
b Abnormal serum Multiple myelomatosis
Low albumin
Myeloma protein
Low γ globulin
c Abnormal serum Nephrosis
Low albumin
High α_2 globulin
Increased γ globulin

Fig. 5, 7

Above.—Diagrammatic representation of apparatus for paper electrophoresis showing paper held between two glass plates with the projecting ends dipping into the electrode vessels.

Below.—Records of paper electrophoresis of sera from three persons: (*a*) normal, (*b*) from a patient with multiple myelomatosis (a blood disorder) showing the altered distribution of proteins, and (*c*) from a patient with nephrosis (a metabolic disorder affecting the kidneys) showing low albumin but increased globulins.

All runs were made in veronal buffer at pH 8 for 18 hours at 4 V/cm.

CHAPTER 6

NUCLEOTIDES, NUCLEIC ACIDS AND NUCLEOPROTEINS

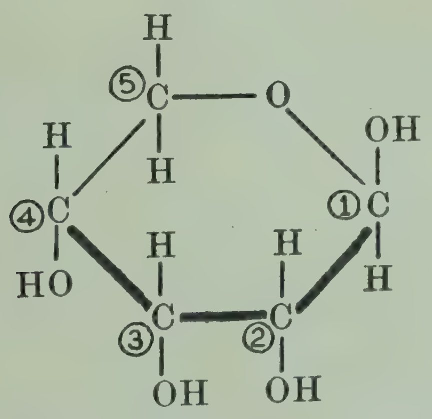
THE nucleoproteins form one of the most important groups of conjugated proteins. They consist of protein conjugated with nucleic acids. On complete hydrolysis the nucleic acids break down to yield purine and pyrimidine bases, pentose sugars and phosphoric acid.

Two main types of nucleic acid are known, the ribonucleic acids (RNA) and the deoxyribonucleic acids (DNA). On hydrolysis RNA yields the pentose sugar ribose, the purine bases adenine and guanine, the pyrimidine bases cytosine and uracil, and phosphoric acid. DNA on hydrolysis yields the sugar deoxyribose, the purine bases adenine and guanine, the pyrimidine bases cytosine and thymine (occasionally also small amounts of 5-methylcytosine and hydroxymethyl cytosine), and phosphoric acid. The two nucleic acids differ therefore in the sugar component and in one of the pyrimidine components as shown in italics in Table 6, 1.

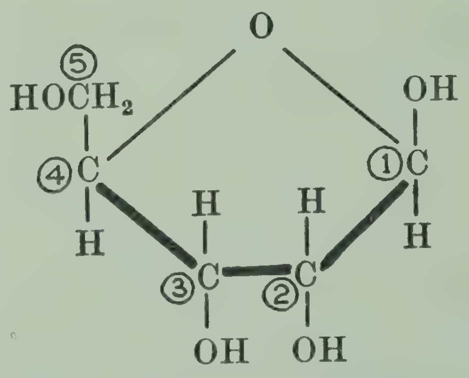
TABLE 6, 1

	Ribonucleic acids	*Deoxyribonucleic acids*
Purine bases	Adenine Guanine	Adenine Guanine
Pyrimidine bases	Cytosine *Uracil*	Cytosine *Thymine* (5-methylcytosine)
Sugar	*Ribose*	*Deoxyribose*
Inorganic component	Phosphoric acid	Phosphoric acid

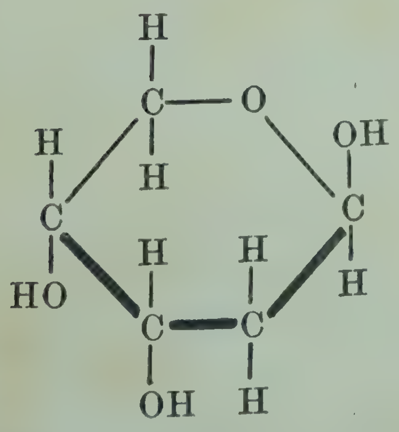
The structures of the sugar components are shown below. In the free state they exist in the pyranose form (see p. 10), but in the combined state in the nucleic acids they exist in the furanose form in which the ring is 5-membered.



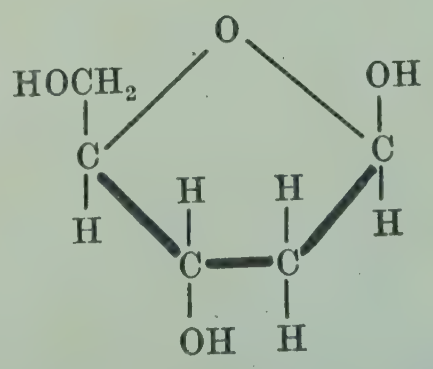
β-D-ribopyranose



β-D-ribofuranose

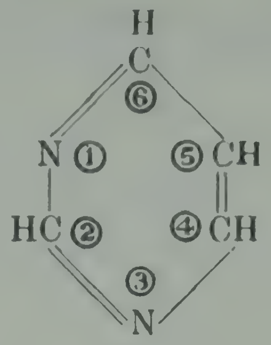


β-D-2-deoxyribopyranose

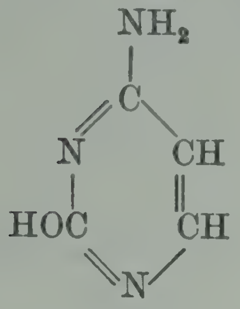


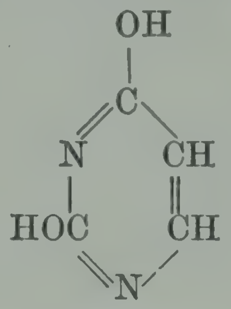
β-D-2-deoxyribofuranose

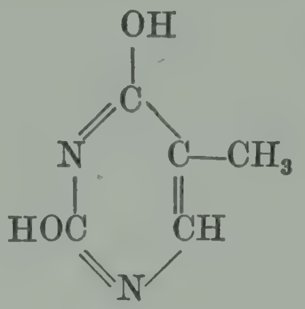
Pyrimidine Bases.—The pyrimidine bases are all derivatives of the parent compound pyrimidine which is structurally a six-membered ring with 2 nitrogen and 4 carbon atoms.

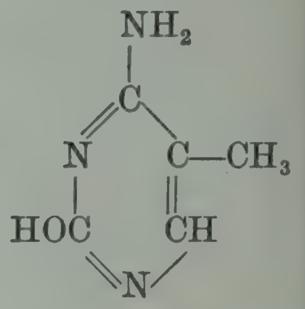


The pyrimidine derivatives found in the nucleic acids are cytosine, uracil, thymine and methyl cytosine.

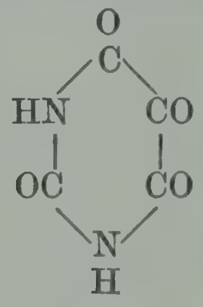

cytosine
2-hydroxy-6-amino-pyrimidine

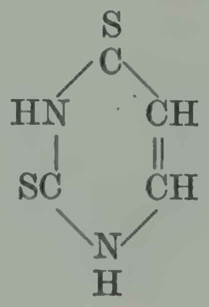

uracil
2:6-dihydroxy-pyrimidine


thymine
5-methyl-uracil

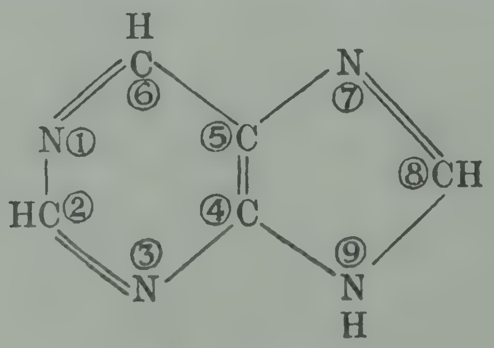

5-methyl cytosine

Two other pyrimidine derivatives are of physiological interest. *Alloxan* when administered to experimental animals can cause diabetes mellitus (p. 307). Derivatives of *thiouracil* are used in the treatment of overactivity of the thyroid gland (hyperthyroidism) (Chapter 51).

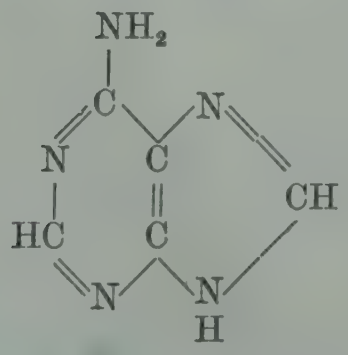

alloxan

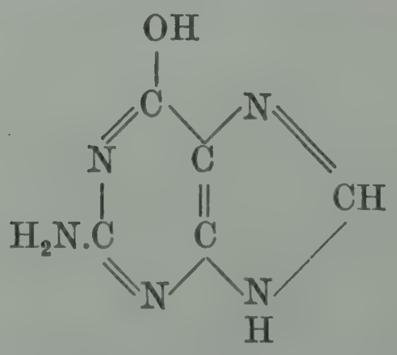

thiouracil

Purine Bases.—The purine bases are derivatives of the parent compound purine which contains a pyrimidine ring and an iminazole ring fused together.

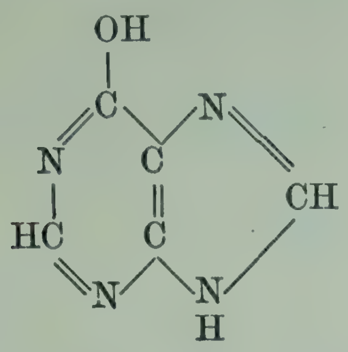

purine

The purine bases which occur in nucleic acids are adenine and guanine.

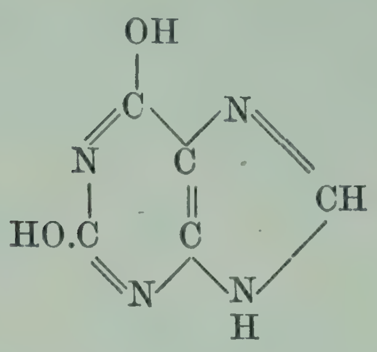

adenine (6-aminopurine)


guanine (2-amino-6-hydroxypurine)

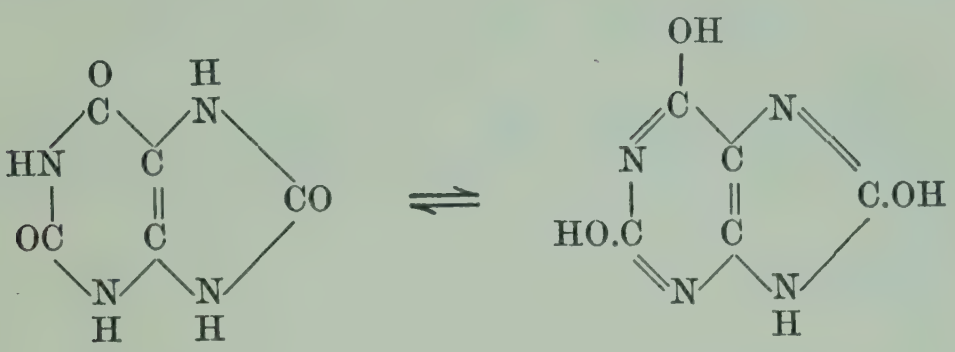
Other naturally occurring purine derivatives are hypoxanthine, xanthine and uric acid.


OH
C
N
N
C
CH
HC
C
N
N
H

hypoxanthine (6-hydroxypurine)


OH
C
N
N
C
CH
HO.C
C
N
N
H

xanthine (2-6-dihydroxypurine)


O
H
C
N
HN
C
CO
OC
C
N
N
H
H
⇌
OH
C
N
N
C
C.OH
HO.C
C
N
N
H

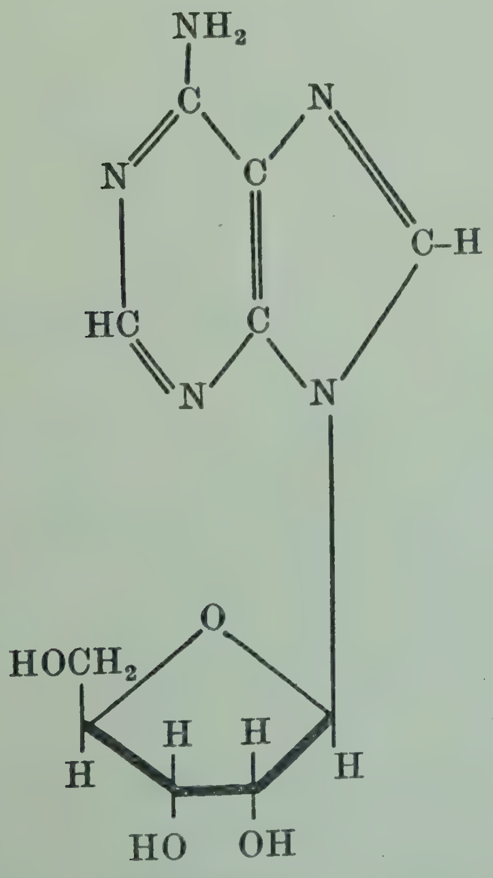
keto form

enol form

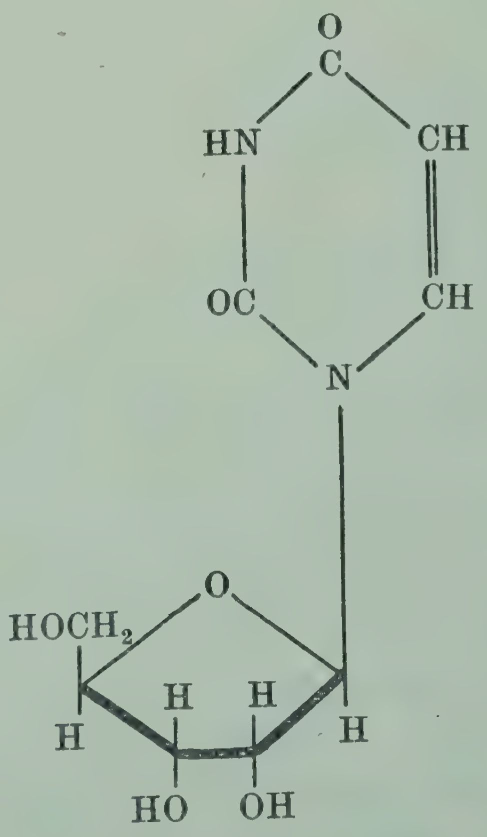
uric acid
(2-6-8-trihydroxypurine)

The purines also include the diuretic drugs *caffeine*, 1 : 3 : 7-trimethyl-xanthine, and *theobromine*, 3 : 7-dimethyl-xanthine.

Nucleosides.—When a purine or a pyrimidine base is condensed with a sugar radical a *nucleoside* is formed. Thus the condensation product of adenine with the pentose ribose is the nucleoside *adenosine*.


NH₂
C
N
N
C
C–H
HC
C
N
N
O
HOCH₂
H
H
H
H
HO
OH

adenosine
(9-β-D-ribofuranosidoadenine)


O
C
HN
CH
OC
CH
N
O
HOCH₂
H
H
H
H
HO
OH

uridine

Similarly guanine on condensation with ribose gives the nucleoside *guanosine*, uracil gives *uridine*, cytosine gives *cytidine*, and so on.

In the nucleosides the nitrogen atom occupying position 3 in the pyrimidine ring or position 9 in the purine ring is linked to carbon atom number 1 (henceforth known as 1′) of the sugar ring.

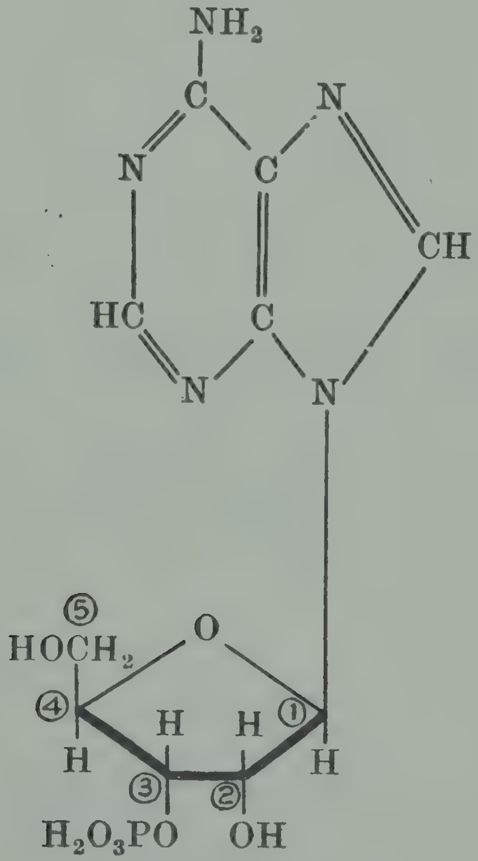
Nucleotides.—When the sugar residue in a nucleoside is condensed with phosphoric acid a *nucleotide* is formed. A nucleotide therefore consists of a base, a pentose sugar and phosphoric acid coupled together,

base—pentose—phosphoric acid.

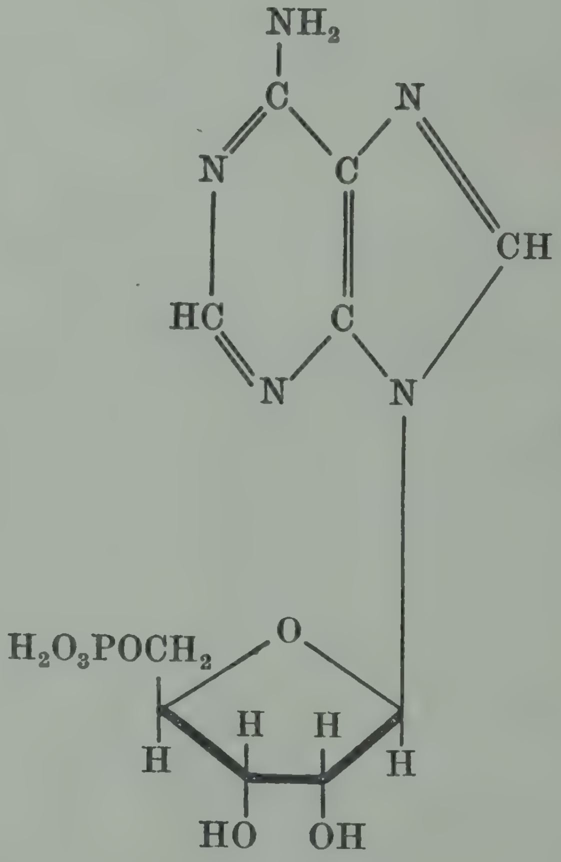
For example, guanosine can condense with phosphoric acid to give the nucleotide *guanylic acid* or *guanosine phosphate*, and similarly adenosine forms a nucleotide *adenylic acid* or *adenosine monophosphate* (AMP),

adenine—ribose—phosphoric acid.

The phosphoric acid residue may be attached in positions 2′, 3′ or 5′ in the ribose ring to form three possible adenine nucleotides, all of which are known.



adenylic acid
(adenosine-3′-phosphate)



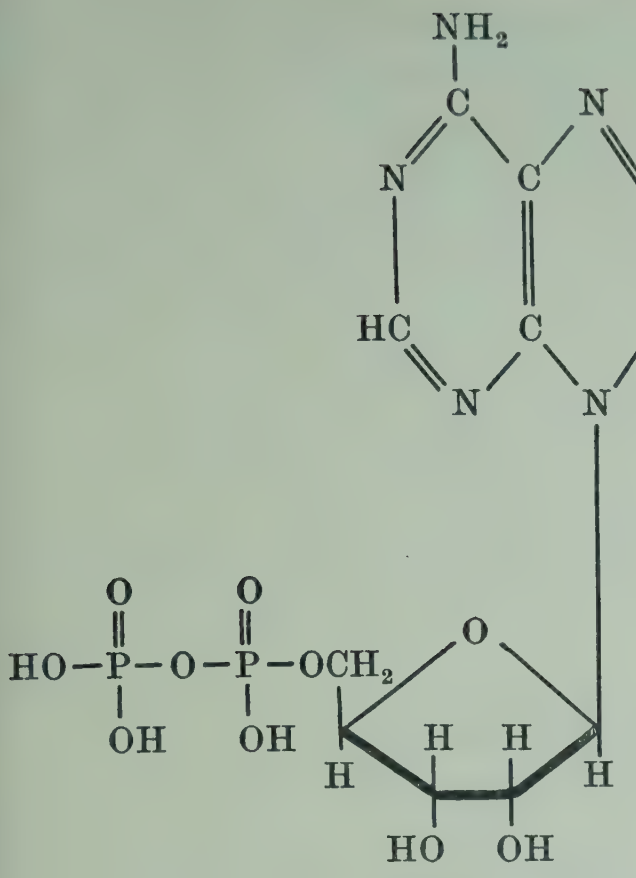
adenylic acid
(adenosine-5′-phosphate)

Nucleosides and nucleotides may also be formed from deoxyribose. Since this sugar has no hydroxyl group in position 2′, only two possible deoxyribonucleotides may be formed, the 3′ and 5′ phosphate derivatives.

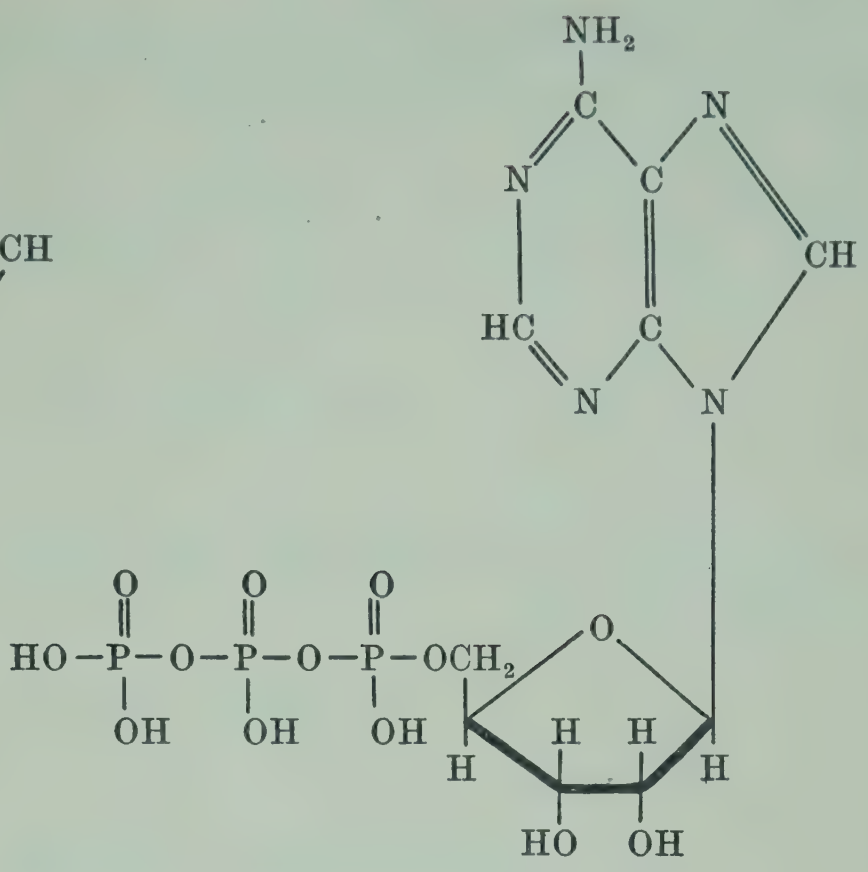
The Nucleotide Coenzymes.—Many important coenzymes (see p. 85) are nucleotides and, although they are not intermediates in the building up of the nucleic acids, their structure may conveniently be discussed at this stage.

Adenosine di- and triphosphates.—Adenosine-5′-phosphate may add on a second phosphoric acid residue to give *adenosine diphosphate* (ADP), which in turn may add on a third phosphate grouping to yield *adenosine triphosphate* (ATP) (or adenyl pyrophosphate).

ATP is one of the most important compounds in the body since the two terminal phosphate groupings are linked by what are known as energy-rich (or high energy) phosphate bonds, sometimes denoted by the symbol ∼.

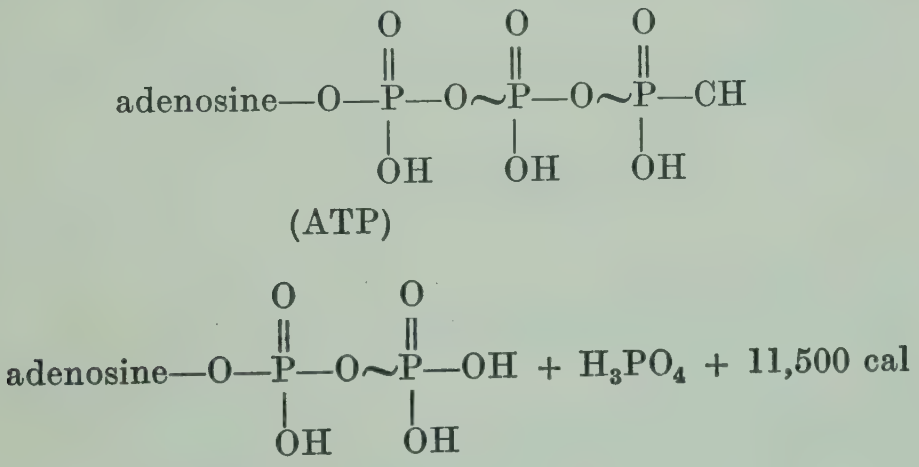


adenosinediphosphate (ADP)



adenosinetriphosphate (ATP)

Liberation of free inorganic phosphate by rupture of these bonds is accompanied by the release of a large amount of energy, about 11,500 cal per mole, whereas the energy released by the rupture of an ordinary phosphate bond such as occurs in adenylic acid or phosphoglycerol is only about 3000 cal per mole.



$$\text{adenosine—O—}\underset{\underset{\text{OH}}{|}}{\overset{\overset{\text{O}}{\|}}{\text{P}}}\text{—O}\sim\underset{\underset{\text{OH}}{|}}{\overset{\overset{\text{O}}{\|}}{\text{P}}}\text{—O}\sim\underset{\underset{\text{OH}}{|}}{\overset{\overset{\text{O}}{\|}}{\text{P}}}\text{—CH}$$

(ATP)

$$\text{adenosine—O—}\underset{\underset{\text{OH}}{|}}{\overset{\overset{\text{O}}{\|}}{\text{P}}}\text{—O}\sim\underset{\underset{\text{OH}}{|}}{\overset{\overset{\text{O}}{\|}}{\text{P}}}\text{—OH} + H_3PO_4 + 11{,}500\ \text{cal}$$

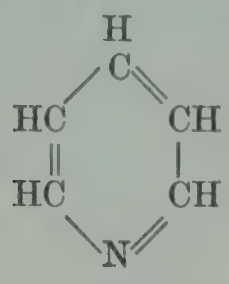
(ADP)

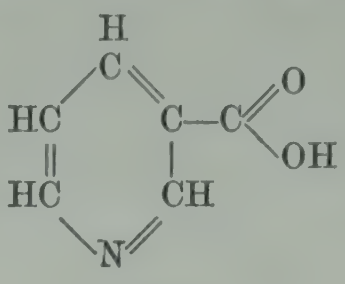
Such energy-rich bonds are also found in creatine phosphate (Chapter 42) and acetyl phosphate, and compounds which contain them are important stores of energy. As we shall see later, ATP is synthesized when carbohydrate is broken down (p. 314) and is used as a source of energy for muscular contraction (Chapter 42).

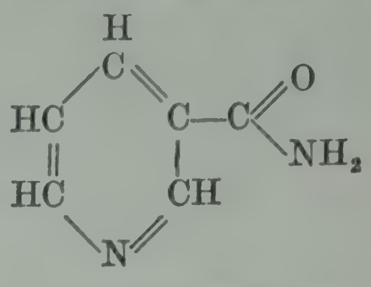
In analogy with AMP, ADP and ATP the nucleosides cytidine, guanosine and uridine from mono-, di- and triphosphates which are designated by the respective contractions CMP, CDP, CTP, GMP, GDP, GTP, UMP, UDP and UTP. All of these compounds have been found as tissue constituents.

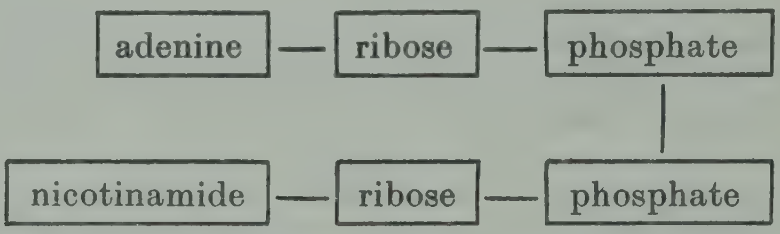
Nicotinamide Nucleotides.—We have seen that the nucleotides which make up nucleic acids contain purine and pyrimidine bases. There are, in addition, certain dinucleotides which because they contain the base *nicotinamide* (*nicotinic acid amide*) are called *nicotinamide* or *pyridine nucleotides.* Nicotinamide is a vitamin of the B complex essential for the prevention of pellagra (p. 127).

One of these nicotinamide nucleotides, known as *coenzyme* I (*Co* I), *diphosphopyridine nucleotide* (DPN), or *cozymase*, contains the two bases adenine and nicotinamide. It consists of adenylic acid condensed through the phosphate grouping with a nucleotide containing nicotinamide, to form a dinucleotide thus:


pyridine

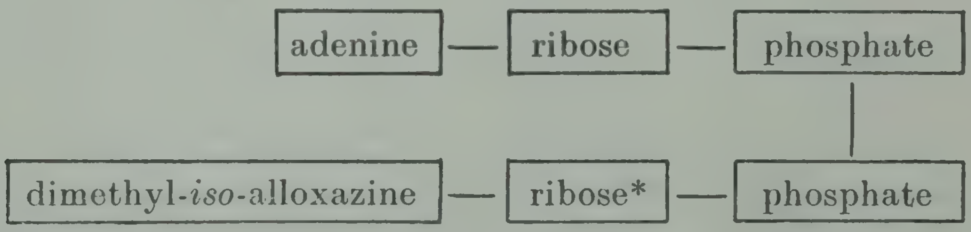

nicotinic acid


nicotinamide
(nicotinic acid amide)


adenine
ribose
phosphate
nicotinamide
ribose
phosphate

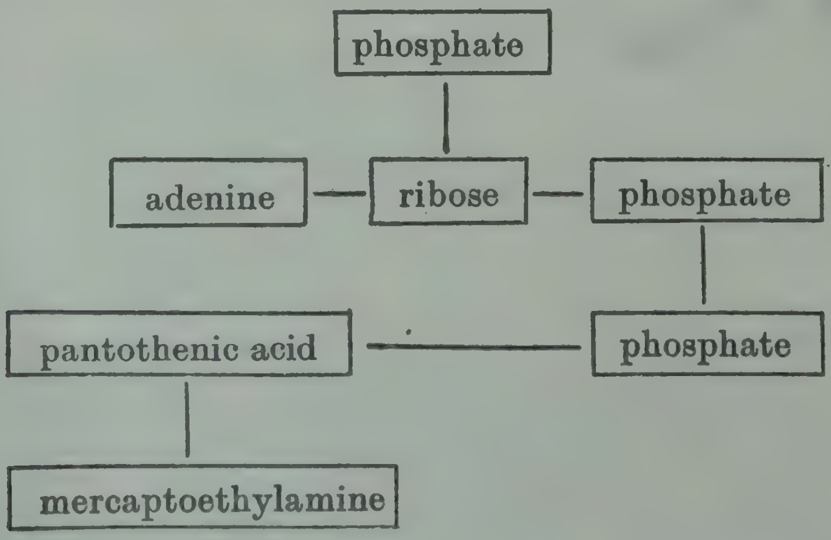
A second such nucleotide is *coenzyme* II (*Co* II), or *triphosphopyridine nucleotide* (TPN), which differs from coenzyme I in containing three, instead of two, phosphoric acid residues in the molecule, the third being attached in position 2′ in the ribose residue which is attached to adenine.

Flavin Nucleotides.—A third important dinucleotide is *alloxazine adenine dinucleotide*, or *flavin adenine dinucleotide* (FAD), which differs from coenzyme I in containing *dimethylisoalloxazine* in place of nicotinamide thus:


adenine
ribose
phosphate
dimethyl-iso-alloxazine
ribose*
phosphate

Both TPN and FAD were discovered by Warburg.

Coenzyme A.—One of the most important coenzymes in carbohydrate and fat metabolism (p. 316 and p. 329) is coenzyme A, discovered by Lipmann. In structure it resembles the pyrimidine nucleotides, but in place of nicotinamide it contains the vitamin pantothenic acid and 2-mercaptoethylamine (thioethanolamine) ($HS—CH_2CH_2—NH_2$):

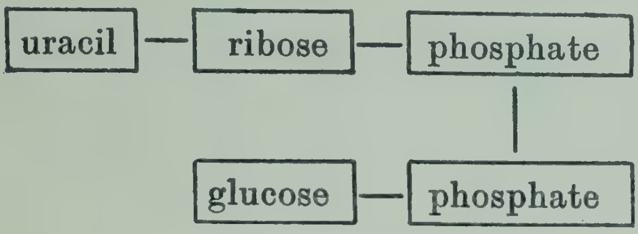

phosphate
adenine
ribose
phosphate
pantothenic acid
phosphate
mercaptoethylamine

---

* More correctly ribitol rather than ribose.

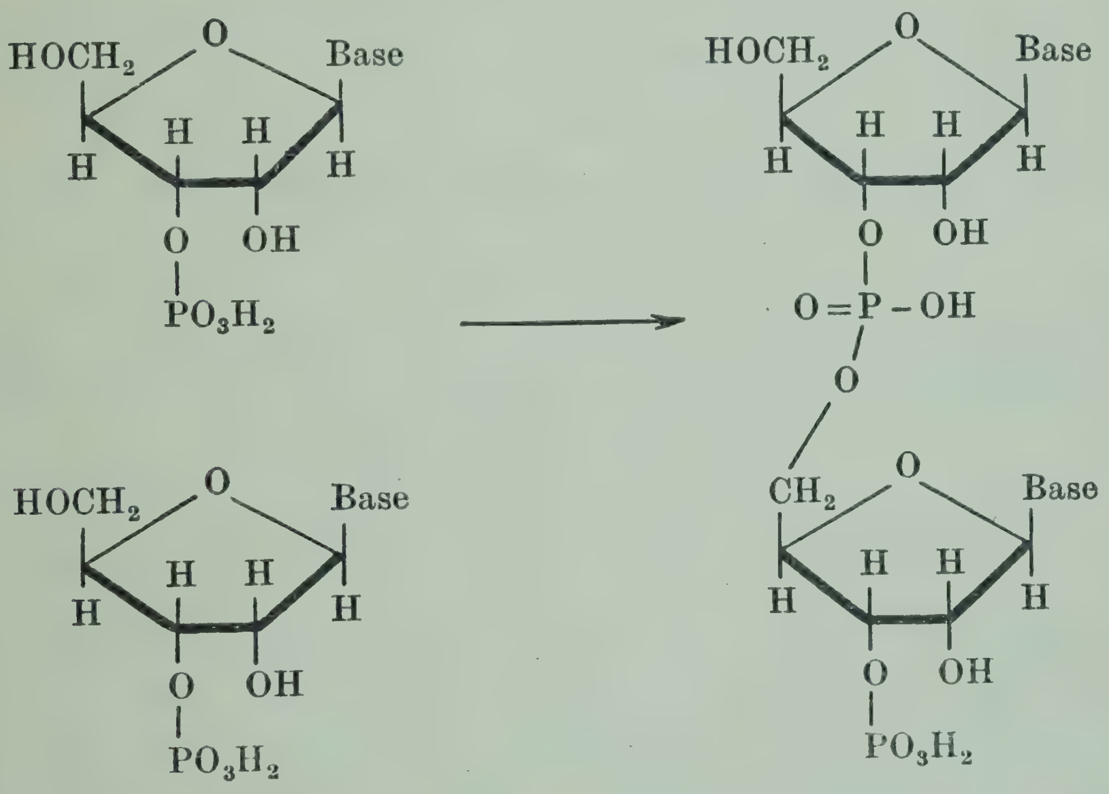
Since the reactive part of the molecule is the sulphydryl (—SH) group, coenzyme A is frequently denoted by the contraction HS—CoA.

Uridine Coenzymes.—The coenzyme nucleotides so far mentioned all contain adenine, but others are known which contain the pyrimidine base uracil. The most important of these is *uridine diphosphate glucose* (UDPG), which is involved in the conversion of galactose to glucose (p. 296).



Polynucleotides, Nucleic Acids

Two nucleotides may condense together with elimination of water to form a dinucleotide in which a phosphate residue links carbon 3′ in one sugar residue with carbon 5′ in the other.

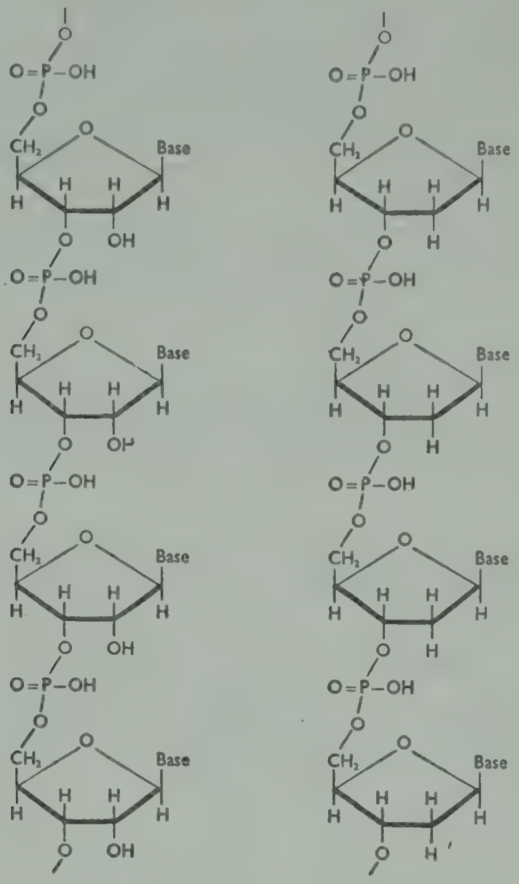


This process may be continued indefinitely until a long chain of nucleotides has been produced forming a polynucleotide. Portions of the chains in RNA and in DNA are shown in Fig. 6, 2.

Although all forms of RNA are built according to the plan shown above, they may vary from each other (*a*) in the molar proportions of bases present (Table 6, 3), and (*b*) in the sequence in which the various bases occur along the polynucleotide chain. It is also possible for branch chains to arise from the main chain at the hydroxyl groups on carbon atom 2′ of the sugar residues.

RNA as extracted from tissues is almost certainly partially degraded, and the molecular weight of the material as it occurs in the native state is high—probably of the order of 20,000. It is easily broken down to a variety of low molecular weight fragments under the influence of the heat stable enzyme *ribonuclease*, which is particularly abundant in pancreas.

In the cell, RNA occurs in combination with protein as *ribonucleoprotein*, chiefly in the cell cytoplasm, but also to a small extent in the nucleus, particularly in the nucleolus. Its exact function is not properly understood but is possibly connected with the process of protein synthesis. It is particularly abundant in cells which are actively engaged in synthesizing protein either for growth or in the process of elaborating enzymes.



Left Right

Fig. 6, 2

Portions of the polynucleotide chains of RNA (left) and DNA (right).

TABLE 6, 3

Molar Proportions of Bases related to Adenine as 10 *in RNA and DNA from different Sources*

Source	Adenine	Guanine	Cytosine	Uracil	Thymine	5-Methyl cytosine
			RNA			
Human liver .	10	16·9	14·7	10·8	—	—
Yeast . .	10	12·1	7·8	9·0	—	—
Tobacco mosaic virus . .	10	9·0	6·1	9·1	—	—
			DNA			
Human thymus.	10	6·4	6·4	—	9·5	0·5
,, liver .	10	6·5	6·6	—	10·0	—
Wheat germ .	10	8·9	6·5	—	10·2	2·2
Tubercle bacillus	10	23·1	23·4	—	9·7	0

Many different forms of DNA are known, all built according to the pattern shown in Fig. 6, 2. No branching of the chain is possible in DNA since there is no hydroxyl group at carbon atom 2′ in the deoxyribose residue.

The different DNAs vary from each other partly in the sequence in which the different bases occur along the polynucleotide chain, and partly in the molar proportions of bases (Table 6, 3). These proportions vary widely from one DNA to another, but are in general fairly constant for the DNA of any one species. In most DNAs adenine and thymine predominate and are present in roughly equimolecular amounts (AT type), but in a few DNAs guanine and cytosine predominate and occur in equimolecular amounts (GC type). In both types the total purines are equivalent to the total pyrimidines, and it has recently been suggested that the DNA molecule consists of two complementary polynucleotide chains arranged facing each other so that the purines in one chain are linked to the pyrimidines of the other chain by hydrogen bonds, forming as it were the rungs of a ladder in which adenine links up with thymine and guanine with cytosine (Fig. 6, 4). If this structure is coiled to form a helix resembling a spiral staircase, we get a crude representation of the modern view of the structure of the DNA molecule as proposed by Watson and Crick (Fig. 6, 5).

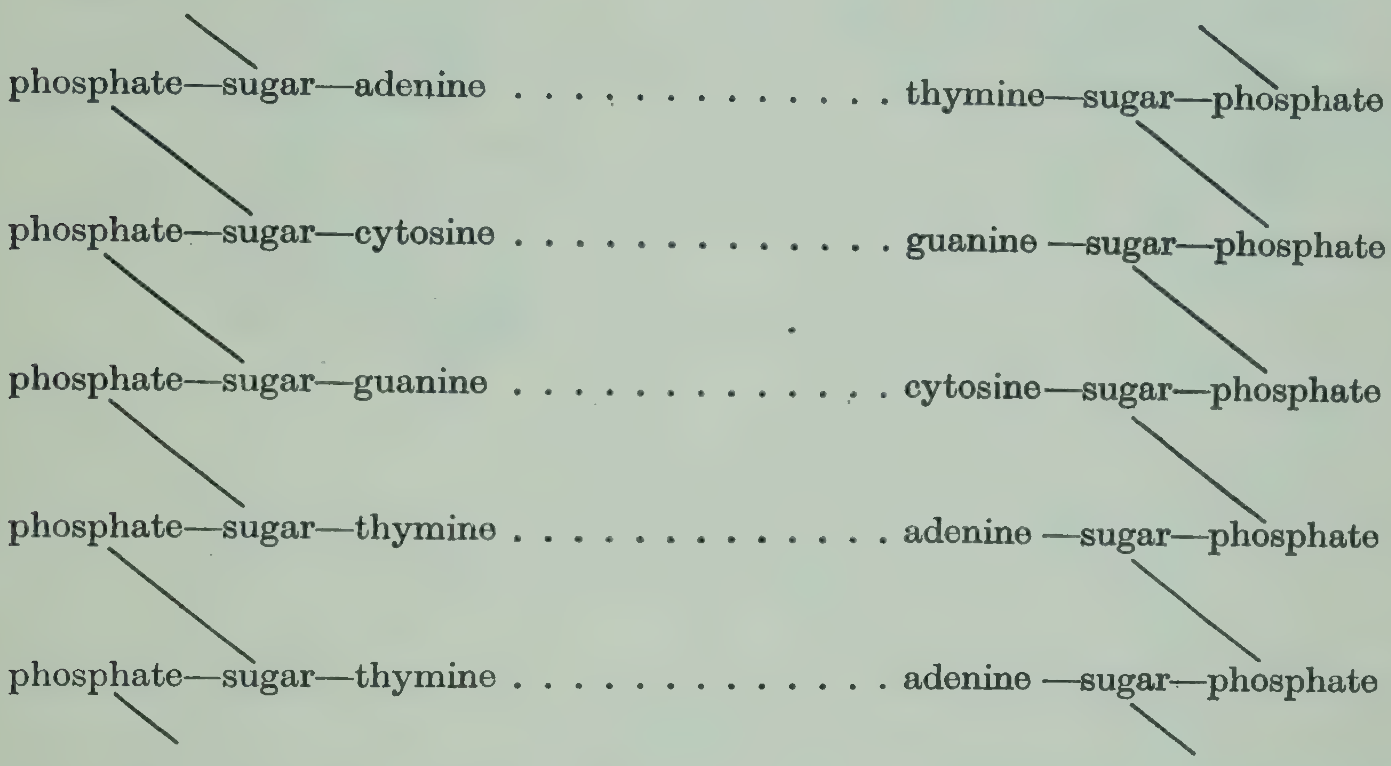

phosphate—sugar—adenine . . . . thymine—sugar—phosphate
phosphate—sugar—cytosine . . . . guanine —sugar—phosphate
phosphate—sugar—guanine . . . . cytosine—sugar—phosphate
phosphate—sugar—thymine . . . . adenine —sugar—phosphate
phosphate—sugar—thymine . . . . adenine —sugar—phosphate

Fig. 6, 4

Diagrammatic representation of part of the polynucleotide chains in DNA.

The molecular weight of DNA is high (about 5,000,000), and the sodium salt of the nucleic acid forms viscous solutions which are attacked by the enzyme *deoxyribonuclease* with breakdown of the polynucleotide chains and consequent loss of viscosity.

DNA is found exclusively in the cell nucleus in combination with histone (p. 54) (or protamine in the sperm heads of certain fish) in the form of *deoxyribonucleoprotein* which is sometimes loosely referred to as *chromatin*. It is an essential component of the chromosomes and plays an important, but as yet not fully understood, part in the mechanism of heredity (Chapter 54).

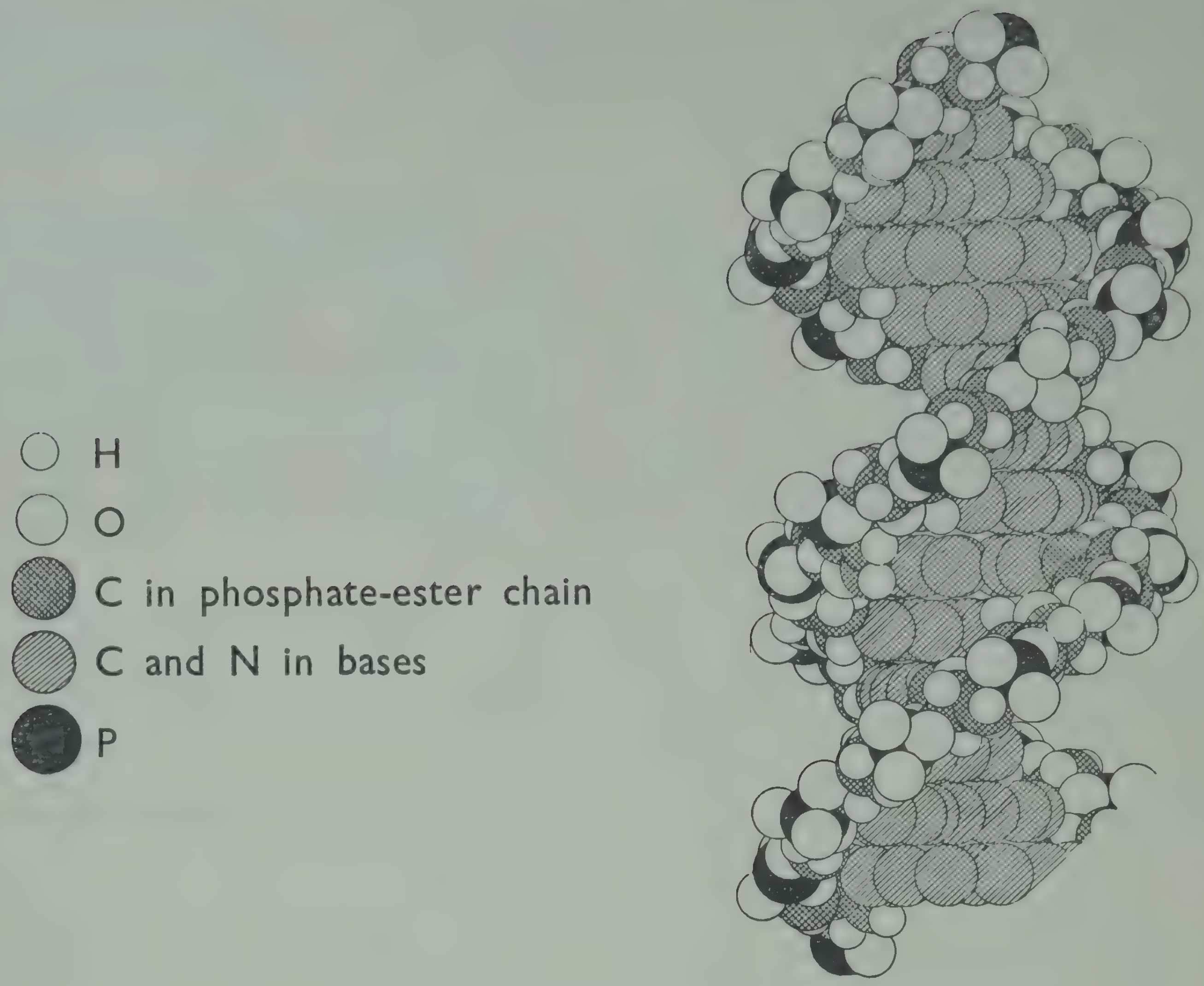


FIG. 6, 5

Diagram of model showing structure of the DNA molecule.
(*By courtesy of M. H. F. Wilkins.*)

CYTOCHEMISTRY

The nucleic acids, being acidic, show an affinity for basic dyes such as toluidine blue, pyronin and methyl green. This affinity is used in selective staining methods for demonstrating the location of nucleic acids in tissue sections. By treating similar sections with ribonuclease or deoxyribonuclease and noting the alterations in staining properties, the position of each type of nucleic acid in the cell may be confirmed.

The nucleic acids moreover, by virtue of their constituent purines and pyrimidines, show an intense and specific absorption of light in the ultraviolet region near 260 mμ. Positive prints of photomicrographs taken in the ultraviolet microscope therefore show dark areas corresponding to the location of the nucleic acids (see, for example, Fig. 22, 2), and the intensity of ultraviolet absorption has been used by Caspersson as the basis of a method for the quantitative estimation of nucleic acids in cells or portions of cells.

By methods of this sort it has been determined that DNA is an exclusively nuclear constituent, whereas RNA is found chiefly in the cytoplasm but also to a small extent in the nucleus, chiefly in the nucleolus.

These results have been confirmed by experiments in which cell constituents have been separated by the process of *differential centrifugation* into their component parts. When cells are disrupted by grinding in a suitable medium such as sucrose solution, and when the cell debris is centrifuged, the nuclei sediment out at low speeds in a short time. Higher speeds bring down particles such as the *mitochondria* of diameter 0·5 to

2·0 μ, while still higher speeds and longer times cause the sedimentation of the *microsomes* which are submicroscopic particles of diameter 60 to 150 mμ. They form the chromophilic ground substance of the cell and, according to one school of thought, are derived from the *endoplasmic reticulum* which forms a fine lace-like pattern of strands in the cell cytoplasm as seen in the electron microscope (Fig. 6, 7). Non-particulate material which does not sediment on centrifuging is said to represent the true hyaloplasm or cell sap.

The localization of these particles in the intact cell is shown in the electron micrograph in Fig. 6, 7.

The distribution of protein and nucleic acid in these cell constituents is shown in Table 6, 6, from which it is clear that most of the protein is in the cell sap and most of the RNA in the microsomes. The enzyme content of these cell fractions has also been determined, and it has become evident that certain enzymes, such as cytochrome oxidase (p. 84) are localized exclusively in the mitochondria, which are the main sites in the cell of oxidation, of energy production and of oxidative phosphorylation (p. 318). They have in fact been termed the 'power plant of the cell.'

TABLE 6, 6

Distribution of Nitrogen Compounds and of Nucleic Acids in Liver Tissue. All figures are percentages of the total in the cell.

	Nitrogen	RNA	DNA
Nuclei . .	15	12	100
Mitochondria	17	4	0
Microsomes .	26	50	0
Cell sap .	42	34	0
Total .	100	100	100

VIRUSES

One of the most interesting groups of nucleoproteins is formed by the filterable viruses. They are important both because many of them have pathogenic properties and because they form a link between the animate and the inanimate worlds, behaving on the one hand as nucleoproteins which can be chemically purified and in some cases even crystallized, and on the other hand as living organisms which can grow and reproduce themselves. They do not, however, respire.

It is well known that the agents which transmit certain infective diseases, for example measles, influenza and yellow fever, are ultramicroscopic particles which can pass through filters capable of holding back bacteria. Such agents are known as filterable viruses and are responsible for a large number of diseases in animals and in plants. Unlike many bacteria, all the viruses are exceedingly difficult to cultivate away from the living tissues of their natural host. They may, however, be made to grow and flourish within the living cells of some foster-host such as the chick embryo. A virus about which much information has been obtained is the tobacco mosaic virus, which is responsible for mosaic disease in tobacco plants. It may be harvested from the leaves of infected plants and can be purified and even crystallized by the usual methods employed in protein purification. The virus after repeated recrystallization has all the chemical properties of a pure nucleoprotein, and yet when it is inoculated into a healthy tobacco plant the virus can grow and reproduce itself and behave as a pathogenic agent causing mosaic disease.

Crystalline tobacco mosaic virus is a nucleoprotein containing ribonucleic acid. It has a very high molecular weight (43,000,000) and exhibits the phenomenon of streaming birefringence. Its molecules appear to be elongated rods which orientate themselves end to end when the solution is stirred. In the electron microscope the molecules are rod-shaped.

The bushy stunt virus which attacks tomato plants has also been crystallized as a ribonucleoprotein (Fig. 6, 8).

While most plant viruses are ribonucleoproteins, the majority of animal viruses such as the influenza virus contain DNA and appear to be structurally more complex than the plant viruses and to be incapable of being crystallized.

REFERENCES

Chargaff, E. & Davidson, J. N. (Edrs.) (1955). *The Nucleic Acids: Chemistry and Biology.* New York: Academic Press.

Davidson, J. N. (1953). *The Biochemistry of the Nucleic Acids*, 2nd ed. London: Methuen.

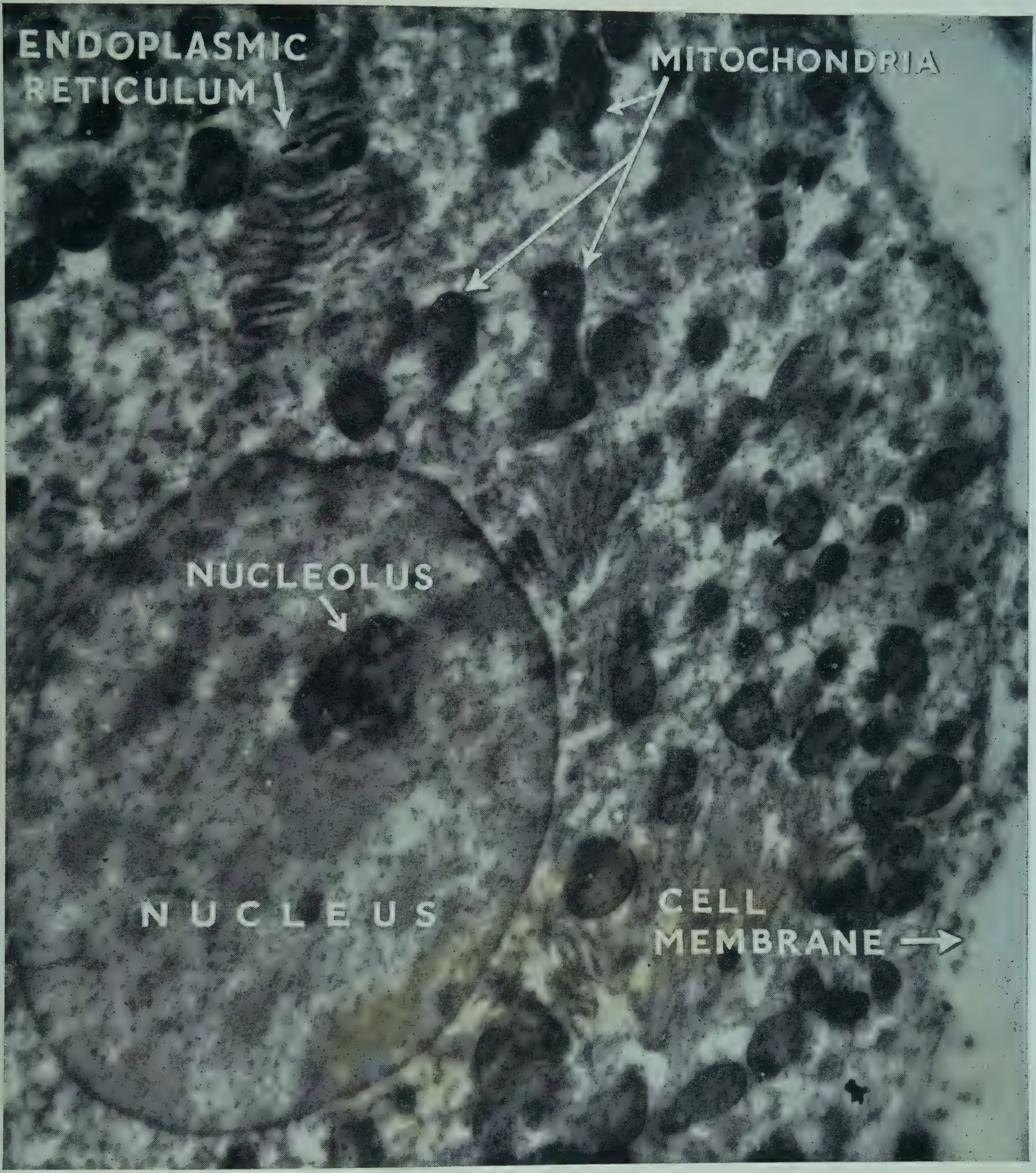

ENDOPLASMIC RETICULUM
MITOCHONDRIA
NUCLEOLUS
NUCLEUS
CELL MEMBRANE

Fig. 6, 7

Electronmicrograph of a portion of a liver cell showing details of the cell structure (Magnification × 8,000

(By courtesy of I. M. Dawson and G. M. Wyburn.)

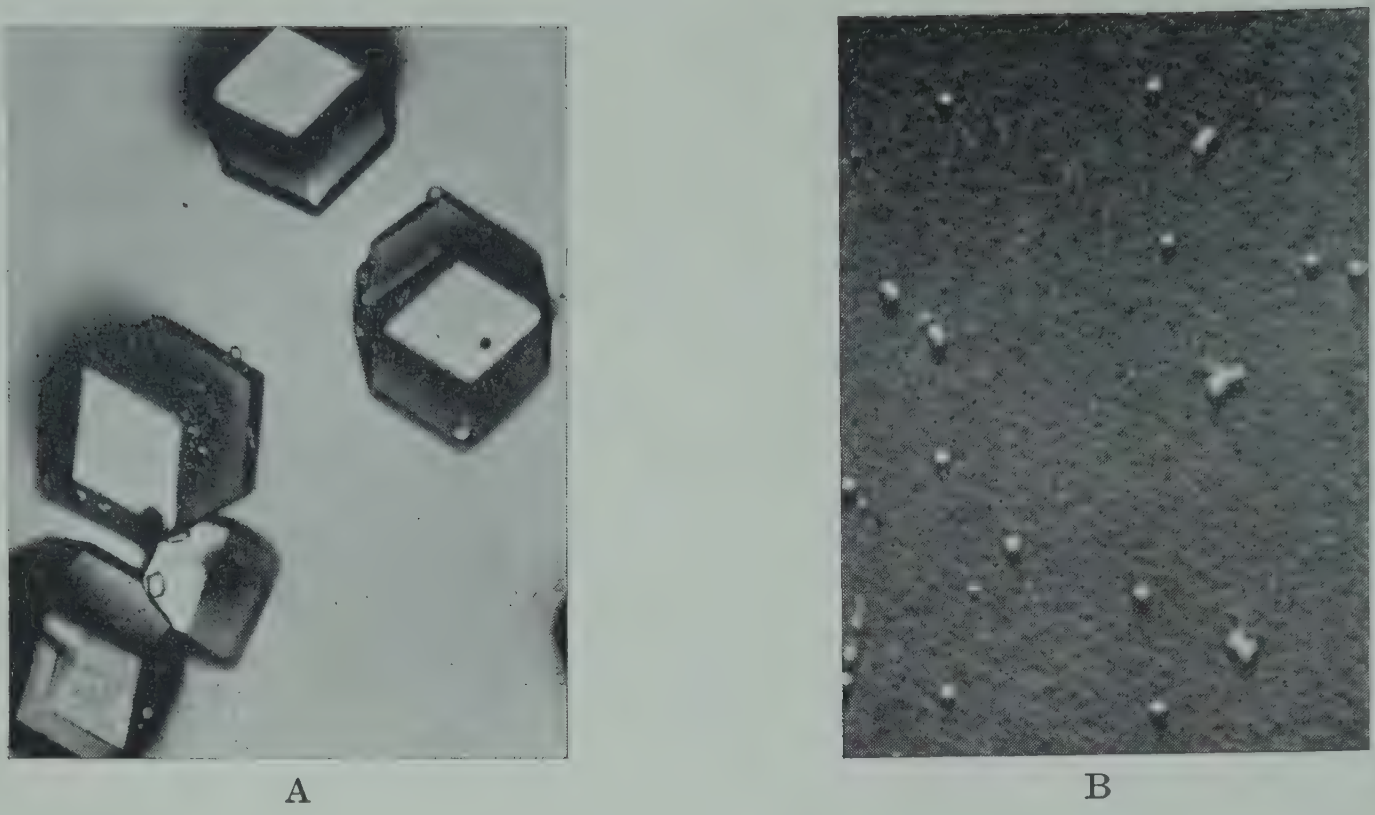


FIG. 6, 8

A. Crystalline bushy stunt virus. × 224.

B. Electron micrograph [gold-shadowed] of bushy stunt virus. × 40,000.

(*By courtesy of W. M. Stanley.*)

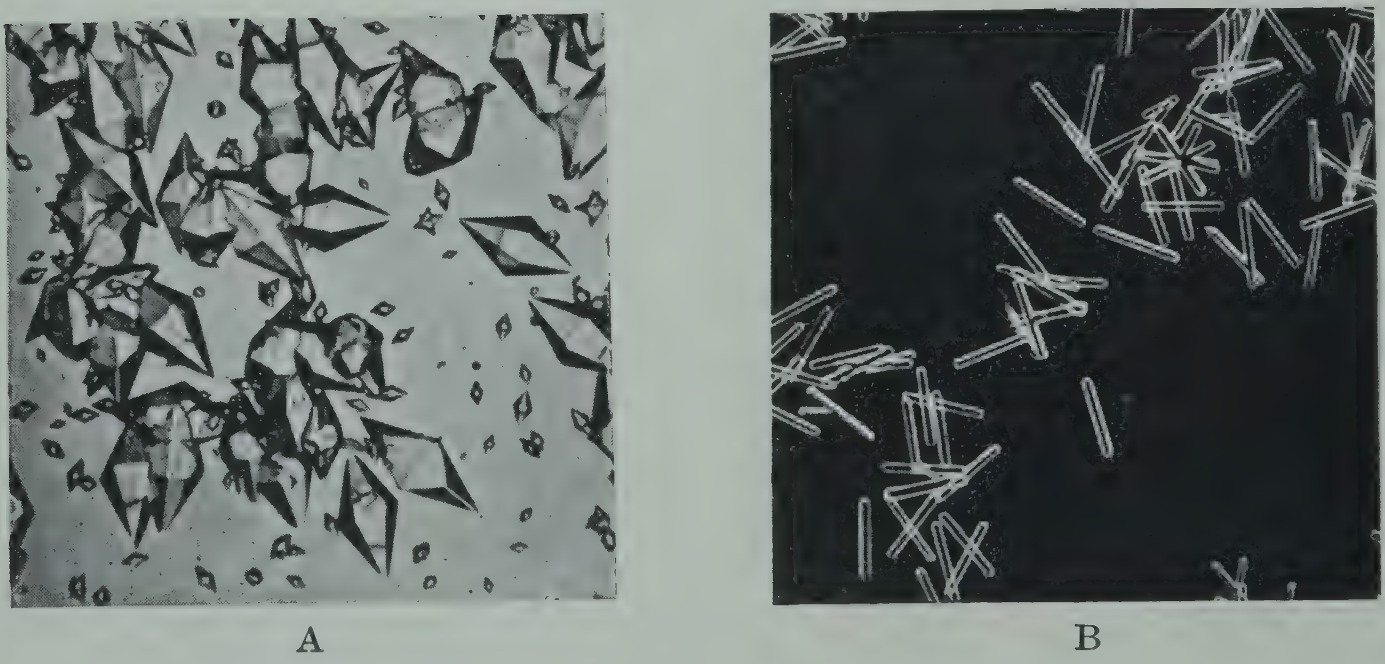


FIG. 7, 1

Photographs of (A) pepsin crystals (× 90) and (B) trypsin crystals (× 202).

(J. H. Northrop (1948), *Crystalline Enzymes*, 2nd ed. Columbia University Press.

CHAPTER 7

ENZYMES

SINCE the time of Berzelius (1825) it has been known that living cells produce substances which act very much in the same way as inorganic catalysts. For such catalysts produced by living cells the name *enzyme* was suggested by Kühne in 1878.

For a time a violent controversy raged between the supporters of Pasteur who held that reactions caused by enzymes, such as alcoholic fermentation, were essentially 'vital' processes which could only be the product of living cells, and the supporters of Liebig who maintained that such reactions could occur without living cells. The controversy was eventually settled by Buchner (1897), who showed that fermentation could be caused by a cell-free extract of yeast. The production of such cell-free enzyme preparations stimulated chemists and physiologists to attempt to isolate enzymes and to determine their chemical nature and this work has culminated in the isolation of a number of enzymes in crystalline form.

Enzymes can be defined simply as catalysts produced by living cells but independent of living cells in their operation. They differ from inorganic catalysts in being rather more delicate; most of them are easily destroyed by heat, they are active over only a limited pH range and are destroyed by acids or alkalies. They are all protein in nature and are characteristically specific in their actions.

Nomenclature.—The substance upon which an enzyme acts is known as the *substrate*. Most enzymes are named after their substrates by adding the suffix '-ase.' Thus *amylase* acts on starch, *uricase* acts on uric acid, *lactase* acts on lactose, and so on. Sometimes more general terms are employed: *proteinases* act on proteins, *lipases* act on fats. Occasionally an enzyme is named after its function rather than after its substrate. Thus *oxidases* catalyse oxidations, and *dehydrogenases* bring about removal of hydrogen. In the case of certain enzymes which have been known for a long time the original names are still used, for example *ptyalin, pepsin, trypsin.*

Intracellular and Extracellular Enzymes.—Although enzymes are catalysts produced by living cells they do not necessarily remain inside the cells which elaborate them. Some cells synthesize enzymes which are immediately cast off. For example the cells of the pancreas produce enzymes necessary for the digestion of food. They are liberated from the cells elaborating them and reach the intestine in the pancreatic juice. Other enzymes remain inside the cells and are concerned with chemical processes there.

In many cases enzymes can be extracted from the cells by grinding them up with quartz sand or by disrupting the cells either in a mechanical homogenizer or with ultrasonic waves, and then extracting the debris with water or buffer solutions. It is only in the case of the enzymes which are soluble in water or in buffers, and which can therefore be extracted, that purification is possible. In other cases the enyzme remains in the cellular debris.

The purification of an enzyme consists essentially in the isolation of one single protein from the complex mixture present in a secretion such as pancreatic juice or in a tissue extract. A large number of enzymes has recently been highly purified and more than twenty have now been obtained in the crystalline state (Fig. 7, 1).

PROPERTIES OF ENZYMES

Since enzymes are proteins they possess many properties common to all proteins. Thus enzymes form colloidal solutions and are of high molecular weight (Table 7, 2).

TABLE 7, 2

Molecular Weights of some Crystalline Enzymes

Enzyme	M.W.	Main Source	Substrate
Ribonuclease	15,000	Pancreas	Ribonucleic acid
Lysozyme	18,000	Egg-white, tears	Glycoproteins
Trypsin	34,000	Pancreas	Protein
Pepsin	36,000	Gastric mucosa	Protein
Chymotrypsin	40,000	Pancreas	Protein
Hexokinase	96,000	Yeast	Glucose
Catalase	248,000	Liver	Hydrogen peroxide
Urease	483,000	Jack bean meal	Urea

These figures should be compared with those for the molecular weights of simple proteins (p. 46).

The catalytic activity of enzymes persists at dilutions so great that chemical tests cannot detect the presence of protein. Thus the presence of the enzyme catalase which decomposes hydrogen peroxide can be demonstrated at a dilution of 1 gram of enzyme in 25,000 litres of water.

The course of enzyme action may be followed by measuring either the rate of disappearance of the substrate or the rate of appearance of the products of the reaction. Thus the progress of the action of the enzyme amylase which catalyses the hydrolysis of starch to maltose may be followed in either of two ways. A sample of the reaction mixture may be removed from time to time and tested with iodine. As the reaction proceeds starch is used up and eventually the blue colour produced on addition of iodine is no longer given. Alternatively the concentration of reducing sugar (maltose) may be estimated from time to time in the reaction mixture. As the reaction proceeds the concentration of maltose increases until all the starch has been broken down. In the case of enzyme reactions which involve the uptake of oxygen, the progress of the reaction may be followed manometrically. Thus the oxidation of uric acid by molecular oxygen under the influence of the enzyme uricase may be allowed to take place in a flask attached to a manometer so that the rate of oxygen utilization can be measured under various conditions.

The Effect of Temperature (Fig. 7, 3).—As the temperature rises, the rate of a chemical reaction increases and, within limits, this is also true of reactions which are catalysed by enzymes. But when the temperature rises above a certain level the protein is denatured. It follows therefore that for every enzyme there is a temperature at which activity is at a maximum, and this is known as the *optimum temperature* (Fig. 7, 3). At the optimum temperature the accelerating effect of rise in temperature is pronounced but heat destruction of the protein is not yet appreciable. For most animal enzymes the optimum temperature lies between 40° and 50°, the point at which destruction of the enzyme begins, while for plant enzymes it lies between 50° and 60°.

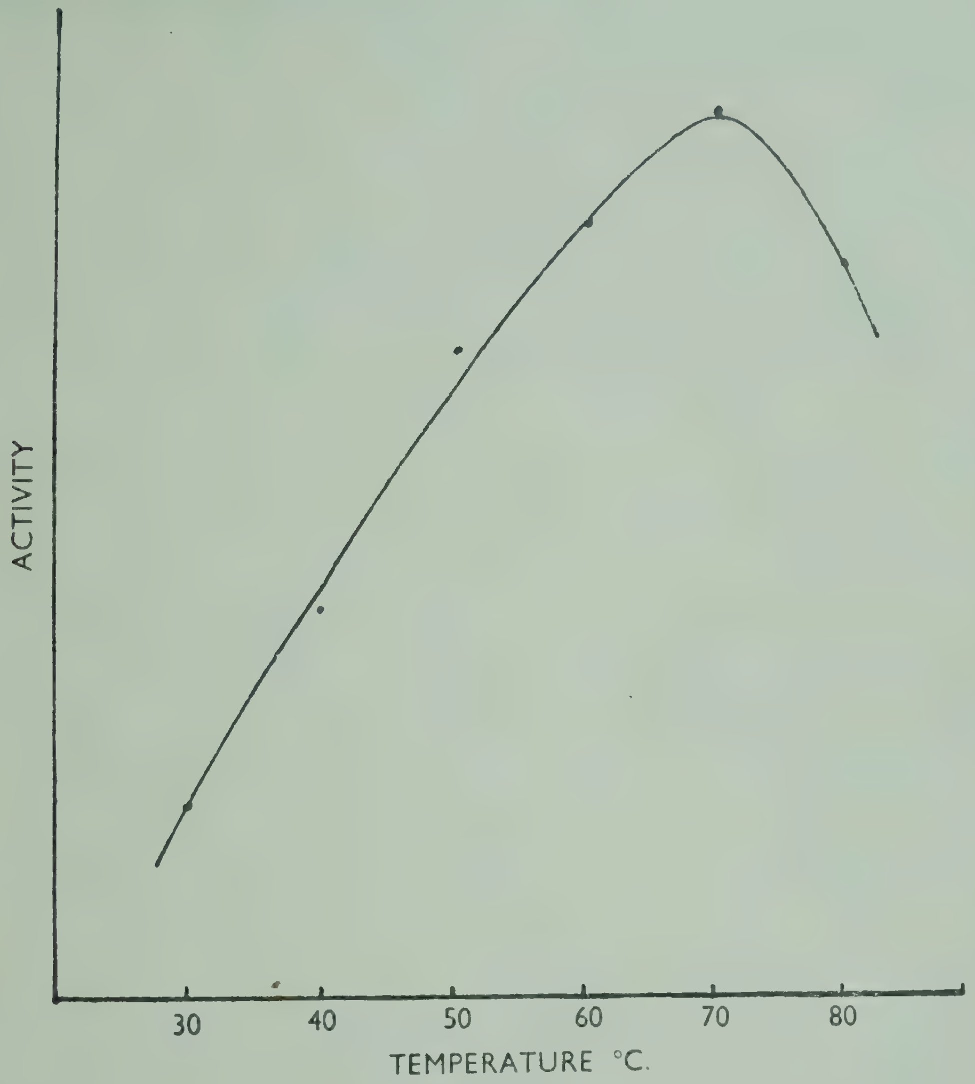


Fig. 7, 3

Effect of change of temperature on the activity of xanthine oxidase.

Most enzymes are completely destroyed at 100° but a few, for example ribonuclease and lecithinase A, can resist boiling in slightly acid solution.

The Effect of pH.—Enzymes are very sensitive to changes in pH and every enzyme functions best within a very limited pH range known as its *optimum pH* (Table 7, 4, and Fig. 7, 5). This value usually lies near

Table 7, 4

Optimum pH Values for some Enzymes

Enzyme	Optimum pH	Substrate
Pepsin	2·7	Protein
Acid phosphatase . .	4·0	Organic phosphates
Ficin (plant) . . .	5·0	Protein
Catalase	6·8	Hydrogen peroxide
Pancreatic amylase . .	6·7–7·2	Starch
Salivary amylase . .	6·8	Starch
Urease	7·0	Urea
Papain (plant) . . .	7·1–7·3	Protein
Ribonuclease . . .	7·6	Ribonucleic acid
Intestinal dipeptidase .	8·0	Dipeptides
Liver esterase . . .	7·8–8·8	Simple esters, *e.g.* ethyl butyrate
Trypsin	8·0–9·0	Protein
Alkaline phosphatase .	8·8–9·1	Organic phosphates
Arginase	9·5–9·9	Arginine

to that of the environment in which the enzyme normally exerts its action in the body. Thus pepsin, which normally acts in the acid medium of the gastric juice, has an optimum pH of 2·7, while intracellular enzymes usually have a pH optimum nearer neutrality.

In some cases the same type of reaction may be catalysed by a number of enzymes with very different optimum pH values. Thus the blood serum normally contains an acid phosphatase (optimum pH 4) and an alkaline phosphatase (optimum pH 9).

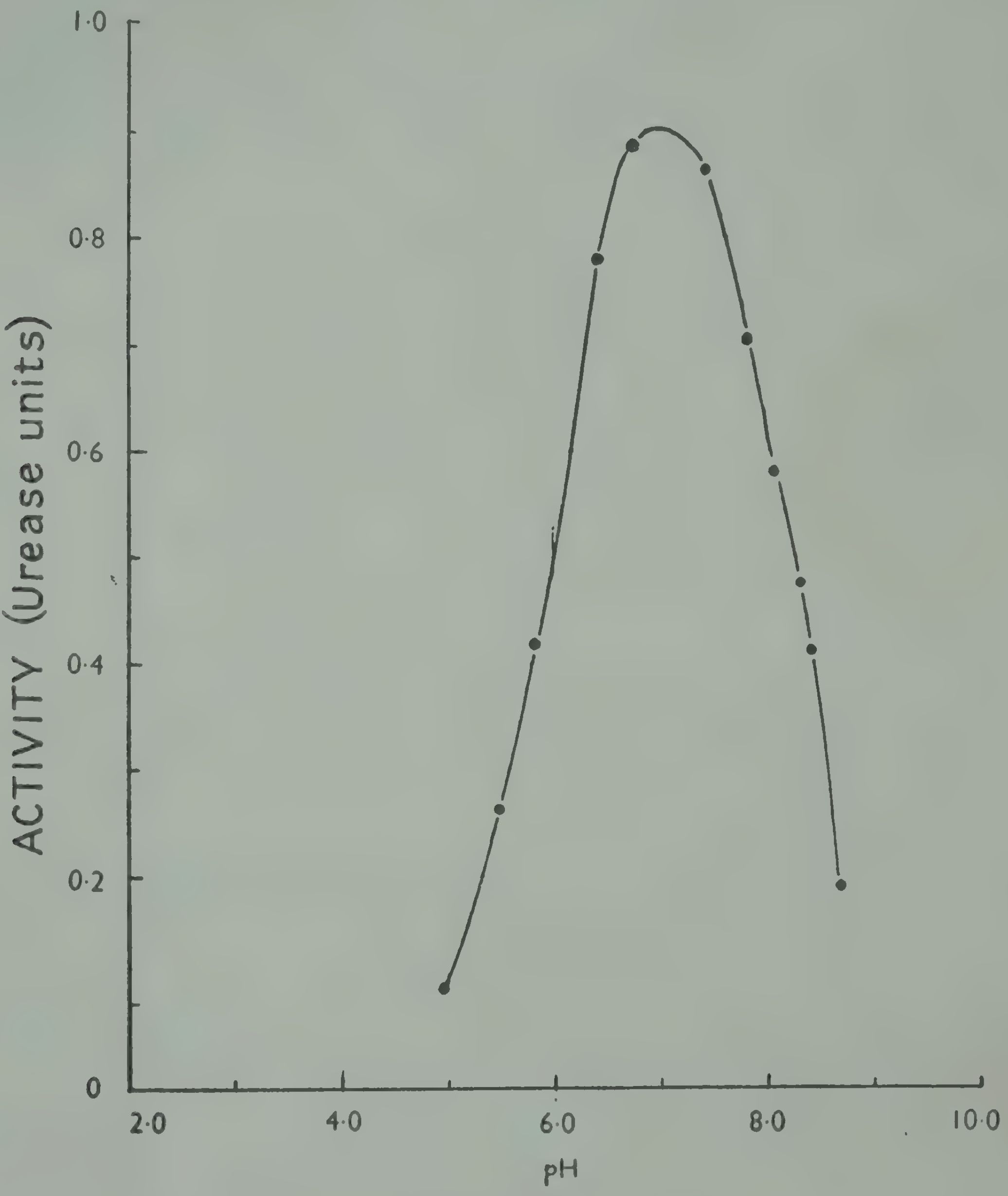


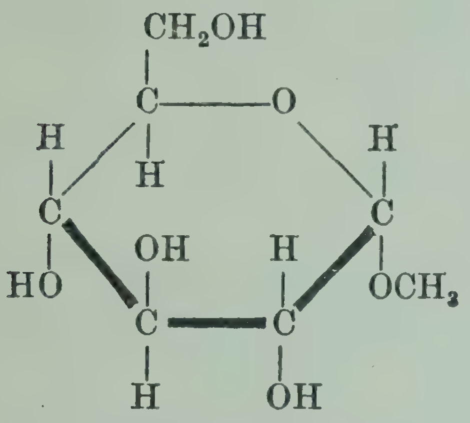
FIG. 7, 5

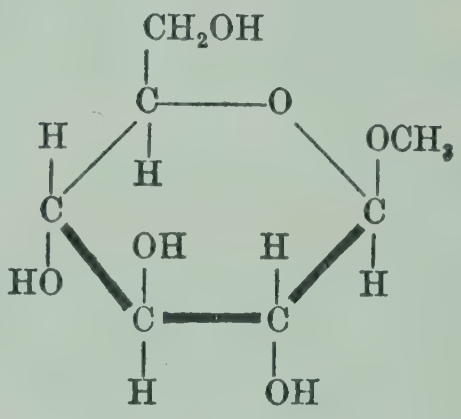
Influence of pH on the activity of urease in phosphate buffer.
(Modified from S. F. Howell & J. B. Sumner (1934), *J. biol. Chem.* **104,** 619.)

The Effect of Concentration of Enzyme and Substrate.—It is usually assumed that a molecule of substrate becomes adsorbed on the surface of the enzyme molecule before the reaction occurs and that the products of the reaction are subsequently cast off, leaving the enzyme in the original state. It follows that one enzyme molecule can take part in the reaction over and over again and that, provided a sufficient excess of substrate is present, the rate of the reaction will within limits be proportional to the concentration of enzyme. In many cases, however, the rate falls off because the enzyme gradually becomes poisoned by the products of the reaction.

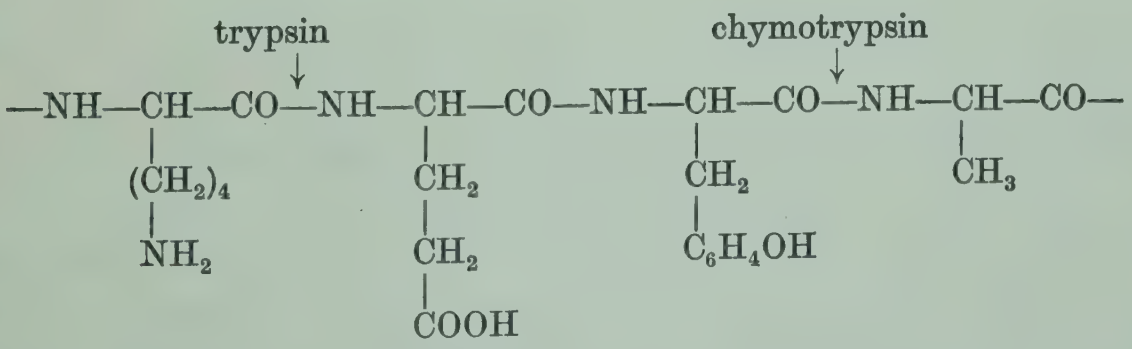
If more than sufficient substrate is present to saturate the enzyme surface, an increase in the substrate concentration does not increase the reaction velocity. Hence in many cases the reaction rate may be constant within rather wide limits of substrate concentration.

Specificity of Enzymes.—Most enzymes act on only one substrate or on one particular group of substrates and have no effect even on closely related compounds. The specificity of enzymes however varies considerably; urease for example acts on urea only, and has no effect upon any other compound, not even upon the substituted ureas. In many cases, however, the action of enzymes is limited not to a single substrate so much as to a group of substances having in common a particular linkage in the molecule. Thus dipeptidase attacks dipeptides in general but not polypeptides, and amylase hydrolyses starch, dextrin and glycogen. Enzyme specificity may, however, be much more restricted, as in the case of maltase which catalyses the hydrolysis of maltose and other α-glucosides only. It can for example catalyse the hydrolysis of α-methyl-glucoside but not of β-methyl-glucoside. The latter can be hydrolysed under the influence of emulsin, which has no effect on α-methyl-glucoside.


α-methyl-glucoside


β-methyl-glucoside

Another illustration of enzyme specificity is to be found among the enzymes which attack the polypeptide chains in the protein molecule. Thus trypsin attacks the bond between the carboxyl group linking lysine or arginine to another amino-acid; while chymotrypsin preferentially acts on the bond between the carboxyl group of an aromatic amino-acid and another amino-acid but is inhibited by any nearby carboxyl group. The action of these enzymes is illustrated in the following diagram:



$$\text{—NH—}\underset{\substack{|\\(CH_2)_4\\|\\NH_2}}{CH}\text{—CO}\overset{\substack{\text{trypsin}\\\downarrow}}{\text{—}}\text{NH—}\underset{\substack{|\\CH_2\\|\\CH_2\\|\\COOH}}{CH}\text{—CO—NH—}\underset{\substack{|\\CH_2\\|\\C_6H_4OH}}{CH}\text{—CO}\overset{\substack{\text{chymotrypsin}\\\downarrow}}{\text{—}}\text{NH—}\underset{\substack{|\\CH_3}}{CH}\text{—CO—}$$

Residues of lysine glutamic acid tyrosine alanine

Inhibitors.—Many enzymes are inhibited by the ions of heavy metals (for example mercury) in low concentration. Those enzymes which are themselves metalloproteins are usually inhibited by cyanides even in minute concentration. Thus cyanide at a concentration of 0·00005 M can completely inhibit the enzyme phenol oxidase which is a metalloprotein. Such inhibition resulting from combination of cyanide with the metal in the enzyme is usually reversible. Thus when phenol oxidase inhibited by potassium cyanide is dialysed against distilled water, the cyanide is removed and full activity is restored to the enzyme.

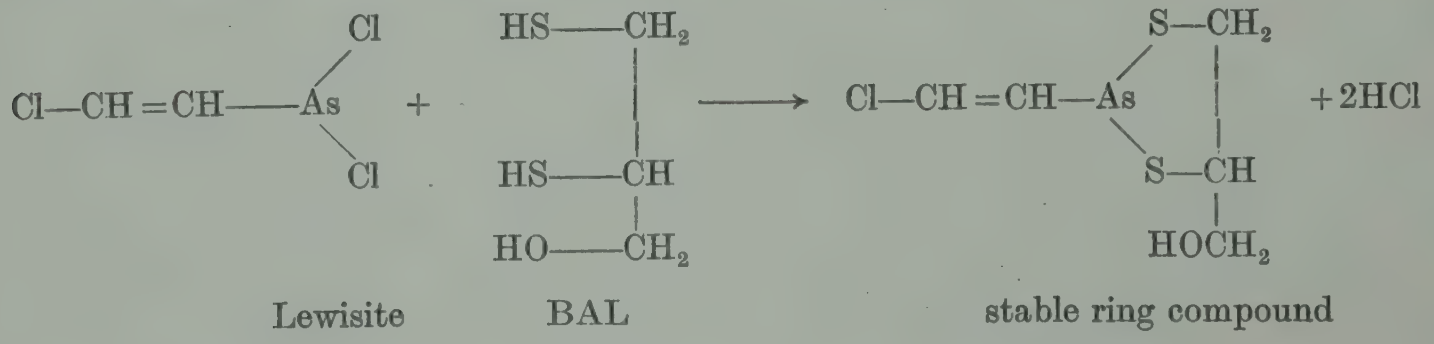
Another example of inhibition is to be found in the case of the enzyme choline esterase which catalyses the hydrolysis of acetyl choline (Chapter 41). This enzyme is normally found in the tissues in amounts which are sufficient to destroy acetyl choline very quickly after it has been liberated. Choline esterase is inhibited by the drug *eserine* (*physostigmine*) which has therefore the effect in the body of prolonging the action of acetyl choline.

It is also powerfully inhibited by such organic derivations of fluorine as diisopropylfluorophosphonate (DFP) which are known collectively as the 'nerve gases.'

Several important enzymes, for example phosphoglyceraldehyde dehydrogenase, are dependent for their activity on the presence in the enzyme molecule of free sulphydryl (–SH) groups to which the substrate becomes attached. Such enzymes are inhibited by iodoacetic acid (ICH_2COOH) which reacts with the –SH groups thus:

$$\text{Enzyme—SH} + ICH_2COOH \longrightarrow \text{Enzyme—S}-CH_2COOH + HI$$

The arsenical war gas lewisite, $ClCH=CH-AsCl_2$, also reacts with free –SH groups and therefore inhibits the enzymes containing them. The action of this poison can be reversed, as Peters has shown, by the use of 2:3-dimercaptopropanol (British anti-lewisite or BAL), a compound with two –SH groups, which reacts with lewisite and reverses its toxic action on the –SH groups of the enzymes of the skin and other tissues.

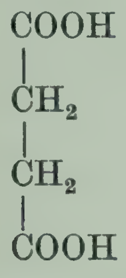


BAL is of value in clinical medicine in the treatment of the toxic effects produced by the administration of certain drugs containing arsenic, gold or mercury.

Competitive Inhibition and Metabolite Antagonists.—The first step in the action of an enzyme on its substrate is the adsorption of the substrate at a specific point on the enzyme surface, where the molecular configuration is such that only the substrate molecule can be neatly accommodated. The reaction then takes place and the products are cast off, leaving the enzyme surface ready to accept a new substrate molecule. To modify the classical simile of Emil Fischer, the substrate fits the enzyme as a key fits the lock.

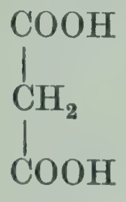
A substance closely similar in structure to the substrate but not identical with it may be acceptable to the enzyme surface at the site normally occupied by the substrate; but since it is not the true substrate it will not be acted on by the enzyme but will remain on the enzyme surface, blocking it and preventing the access of true substrate molecules. Such substances are known as *competitive inhibitors* since they compete with the substrate for access to the appropriate part of the enzyme surface. To extend the classical simile, we may say that a competitive inhibitor acts as a key which can enter the lock but jams instead of turning it.

A good example of competitive inhibition is to be found in the case of the enzyme succinic dehydrogenase (succinoxidase) which oxidizes succinic acid. Malonic acid has a structure so similar to that of succinic acid that it is adsorbed on the enzyme surface although it is not acted on further by the enzyme. It therefore acts as a competitive inhibitor, its molecules jostling with those of the substrate for the place of attachment on the enzyme surface. Malic and oxaloacetic acids likewise act as competitive inhibitors. On the other hand, the enzyme is non-competitively inhibited by iodoacetic acid since it contains –SH groups.



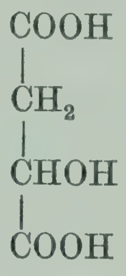
$$\begin{array}{c}\mathrm{COOH}\\|\\\mathrm{CH_2}\\|\\\mathrm{CH_2}\\|\\\mathrm{COOH}\end{array}$$

succinic acid



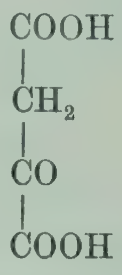
$$\begin{array}{c}\mathrm{COOH}\\|\\\mathrm{CH_2}\\|\\\mathrm{COOH}\end{array}$$

malonic acid



$$\begin{array}{c}\mathrm{COOH}\\|\\\mathrm{CH_2}\\|\\\mathrm{CHOH}\\|\\\mathrm{COOH}\end{array}$$

malic acid



$$\begin{array}{c}\mathrm{COOH}\\|\\\mathrm{CH_2}\\|\\\mathrm{CO}\\|\\\mathrm{COOH}\end{array}$$

oxaloacetic acid

When enzyme inhibition is non-competitive the concentration of inhibitor necessary to produce a given effect is independent of the substrate concentration, whereas in competitive inhibition the degree of inhibition depends on the ratio of the concentration of inhibitor to that of substrate. It follows that if the concentration of inhibitor is kept constant, the inhibition can be progressively overcome as the concentration of substrate is raised.

The principle of blocking a biochemical reaction by the use of a compound similar to but not identical with the natural participant in the reaction is of great importance in biology, and is illustrated again on p. 134 in relation to anti-vitamins and chemotherapy.

Activators.—The term *activator* is usually employed to indicate a substance which increases the activity of an enzyme in a non-specific manner. Thus the presence of chloride ions accelerates the action of amylase and the presence of sulphydryl compounds (containing the –SH group) or of cyanides accelerates the action of the plant protease papain, although neither substance is essential for the activity of the enzyme. Activators are not essential for the action of an enzyme which can act, although more slowly, without them. They must be distinguished clearly from coenzymes which are specific and essential components of an enzyme system; in the absence of the coenzyme the enzyme system is inactive.

Some activators act merely by removing or neutralizing traces of an inhibitory factor such as a heavy metal but the mode of action of others is much more obscure.

Anti-enzymes.—Certain tissues contain protective substances called *anti-enzymes* which inhibit specifically the action of enzymes. Thus the cells of the intestinal mucosa contain anti-enzymes which protect them from digestion by the proteolytic enzymes engaged in attacking the proteins of the food. The round worm, *Ascaris*, contains anti-enzymes which prevent its being digested by the proteolytic enzymes, pepsin, rennin and trypsin, which it normally encounters in the alimentary canal, but it is quickly digested alive by the proteinase *ficin* of the fig-tree which in ordinary circumstances the worm does not encounter.

When an animal is given repeated small injections of a pure enzyme its blood serum may subsequently be found to contain a specific anti-enzyme. Thus the serum of rabbits which have received repeated minute injections of crystalline urease contains an anti-urease which can combine with urease and inhibit its action. The enzymes in snake-venom can similarly cause the production of anti-enzymes (antivenins) when injected into animals in small doses, and such antivenins are used therapeutically to neutralize the toxic effects of snake-bite.

Zymogens and Kinases.—Enzymes sometimes exist in the form of inactive precursors called *zymogens* which are activated on coming into contact with a *kinase*. Thus the pancreas and pancreatic juice contain the zymogen *trypsinogen*, which is itself inactive but which is converted into active *trypsin* on coming into contact with previously formed trypsin or with the kinase, *enterokinase*, of the duodenal mucosa.

By producing zymogens the organism protects itself from the action of active enzymes in situations where their activity might prove injurious.

THE STRUCTURE OF ENZYMES

It is generally believed that enzymes owe their activity to some active group on the surface of the protein molecule. The substrate is adsorbed on the surface close to this active group and reacts with it. The rest of the protein acts as a carrier for the active group but may also facilitate its reaction with the substrate in some way in addition to bringing the two into spatial proximity.

This active group may in some cases, for example in pepsin, be simply a special arrangement of amino-acids in one part of the protein molecule but in other cases it is a special grouping of non-protein nature and in these cases it is called a *prosthetic group*. An example of this latter type is peroxidase in which the prosthetic group is an iron-porphyrin complex similar to the haem of haemoglobin (p. 410). The prosthetic group may also consist of a metal atom combined with the protein, for example copper in phenol oxidase and zinc in carbonic anhydrase.

Usually a prosthetic group is firmly attached to the protein but in some cases the non-protein part of an enzyme is readily removed and forms a dissociable complex with the protein. In this case the non-protein portion is referred to as a *coenzyme*. The protein part is then called the *apoenzyme* and the active complex the *holoenzyme*.

Structurally enzymes may be divided into three main groups:

1. Enzymes which consist exclusively of protein (for example pepsin).

2. Enzymes consisting of protein plus some inorganic component such as a metallic ion (metalloprotein) (for example tyrosinase). The metal ion may serve merely as a point of attachment of the substrate to the protein. In other cases it may play a more active part, undergoing a continuous change of valency by electron transfer, as in the case of the enzymes containing iron (for example peroxidase). In such enzymes, as we shall see later, the iron is not attached directly to the protein but is built into a complicated porphyrin molecule, which in turn is bound to the protein.

Apart from iron, other metals are found in certain enzymes. Thus purified phenol oxidase contains 0·2 per cent., purified ascorbic acid oxidase 0·25 per cent., and purified tyrosinase 0·25 per cent. of copper. Purified carbonic anhydrase contains 0·31 to 0·34 per cent. of zinc, arginase contains manganese and carboxylase contains magnesium.

3. Enzymes which consist of protein plus a non-protein organic compound of low molecular weight (a *coenzyme*) bound to it more or less loosely. A metallic ion may be present as well, for example in carboxylase.

COENZYMES AND APOENZYMES

We have already seen that some enzymes appear to consist of a simple protein alone, while in other cases the enzyme requires for its activity the presence of a *coenzyme*. Such coenzymes are best regarded as prosthetic groups which can be split off and isolated as separate substances. The protein part of the enzyme alone is known as the *apoenzyme*. Where the coenzyme is only loosely attached to the apoenzyme it can usually be separated by dialysis, or by precipitation of the protein with ammonium sulphate at some suitable pH value, leaving the coenzyme in solution.

Many coenzymes are complex derivatives of vitamins of the B group (Chapter 10), and in many coenzymes the vitamin is built up into a complex nucleotide-like structure such as the pyridine nucleotides (nicotinamide nucleotides) and flavin-adenine dinucleotide (FAD) which have already been discussed on p. 61 and p. 62. Such coenzymes play a very important part, as we shall see later, in the processes of biological oxidation in which, in association with the appropriate apoenzyme, they participate by accepting, and subsequently giving up, pairs of hydrogen atoms.

It should be noted that one and the same coenzyme may react with different apoenzymes to yield different enzyme systems. For example the vitamin thiamine (vitamin B_1) is phosphorylated in the tissues to yield diphosphothiamine (sometimes also known as cocarboxylase), which can act as coenzyme in at least four different enzyme systems in each of which it combines with a different apoenzyme.

REVERSIBILITY OF ENZYME-CATALYSED REACTIONS

Since enzymes are catalysts they can accelerate the attainment of equilibrium in reversible reactions by catalysing both the forward and the reverse reactions, and we might expect, therefore, that one and the same enzyme system would catalyse both the breakdown of a substance and its synthesis. This process can be demonstrated very conveniently in the case of the enzyme phosphorylase (p. 294), which catalyses the reaction

$$\text{glycogen} + H_3PO_4 \xrightleftharpoons{\textit{phosphorylase}} \text{1-phosphoglucose}$$

By suitable variations of the conditions the enzyme can be used either for the preparation of 1-phosphoglucose or for the *in vitro* synthesis of glycogen. The point of equilibrium in such enzyme-catalysed reactions depends among other things upon the concentrations of reacting substances.

Urease, for example, is usually associated with the hydrolysis of urea to ammonium carbonate,

$$\text{urea} + 2H_2O \xrightleftharpoons{\textit{urease}} \text{ammonium carbonate}$$

but by incubating ammonium carbonate solution with urease, under conditions in which urea is removed as fast as it is produced, for example by continuous extraction with butanol, the synthesis of urea can be demonstrated.

THE LOCALIZATION OF ENZYMES IN THE CELL

It is now possible to determine in which part of the cell a given enzyme is located. This may be done by either of two methods. In the *histochemical* procedure a suitably prepared section of tissue is incubated with the substrate under rigidly controlled conditions which are so arranged that the products of the reaction are revealed by their interaction with some suitable reagent to form a coloured product. The sites of greatest enzyme activity in the cell thus appear most deeply stained.

The second approach is to separate the various cell constituents in bulk by the usual cytochemical procedures outlined in Chapter 6, and to determine the enzyme content of each fraction.

While some enzymes are present in many or all cell fractions, a few are exclusively localized in one or other cell component. Thus, cytochrome oxidase (p. 84) is found exclusively in the *mitochondria* along with the fatty acid oxidases and most, but not all, of the enzymes of the citric acid cycle (Chapter 20). The process of oxidative phosphorylation (p. 318) involving ATP synthesis occurs only in the mitochondria. The *microsomes* contain nearly all the esterases of the cell while the *cell sap* contains many of the enzymes of glycolysis (p. 311) together with the two enzymes of the citric acid cycle (aconitase and isocitric dehydrogenase) which are missing from the mitochondria.

AUTOLYSIS

If an organ or tissue is removed from the body and stored in aseptic conditions, that is, uncontaminated by bacteria, a slow disintegration and liquefaction of the tissue takes place. This process, which is called *autolysis*, occurs in the body after death and is accelerated by an increase of temperature. It proceeds much more rapidly in glandular tissues such as liver, kidney and pancreas than in muscle and connective tissue. After the death of the cells the enzymes of respiration become inactive, whereas the action of the hydrolytic enzymes persists often at a more rapid rate, so that protein, the major constituent of the cells, is broken down, amino-acids and ammonium salts being found among the end-products. The protein-splitting enzyme responsible for these changes is called *cathepsin*. Autolysis may be prevented or arrested by 'fixation' of the tissue, that is to say by impregnating it with an agent, such as formalin or alcohol, which denatures, and therefore inactivates, the enzyme protein.

REFERENCES

LAIDLER, K. J. (1954). *Introduction to the Chemistry of Enzymes.* New York: McGraw Hill.

McELROY, W. D. & GLASS, B. (Edrs.) (1954). *The Mechanism of Enzyme Action.* Baltimore: Johns Hopkins Press.

NEILANDS, J. B. & STUMPF, P. K. (1955). *Outlines of Enzyme Chemistry.* New York: Wiley.

NORTHROP, J. H. (1948). *Crystalline Enzymes,* 2nd ed. New York: Columbia University Press.

Some aspects of enzyme research (1953). *Brit. med. Bull.* **9**, No. 2.

SUMNER, J. B. & MYRBACK, K. (1950). *The Enzymes.* New York: Academic Press.

WORK, T. S. & WORK, E. (1948). *The Basis of Chemotherapy.* Edinburgh: Oliver & Boyd.

CHAPTER 8

BIOLOGICAL OXIDATIONS AND REDUCTIONS

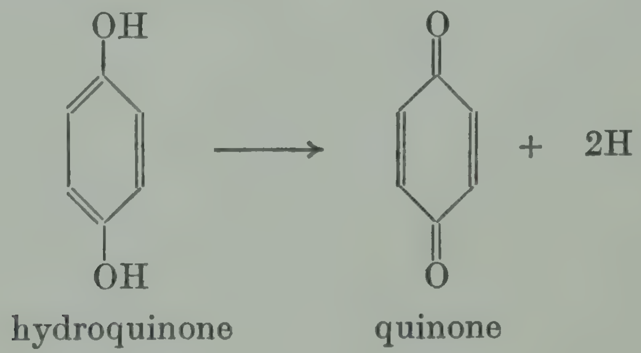
WE shall see later how oxygen is transported by the blood from the lungs to the tissues, where it is utilized to bring about the oxidation of products of metabolism with the liberation of energy. Such oxidations require the influence of powerful enzymes, many of which involve several components. Stearic acid and hypoxanthine, for example, can be oxidized in the laboratory only by powerful oxidizing agents at high temperature; in the body they are oxidized easily and quickly at 37° owing to the presence of suitable enzyme systems.

It should be understood that the process of oxidation may involve any one of four possible types of chemical change:

1. A gain of oxygen:

$$CO + \tfrac{1}{2}O_2 \longrightarrow CO_2$$

2. A loss of hydrogen:



hydroquinone quinone

3. Addition of water with loss of hydrogen:

$$\underset{\text{chloral}}{CCl_3CHO} + H_2O \longrightarrow \underset{\text{chloral hydrate}}{CCl_3CHO.H_2O}$$

$$\underset{\text{chloral hydrate}}{CCl_3CHO.H_2O} \longrightarrow \underset{\text{trichloracetic acid}}{CCl_3COOH} + 2H$$

4. The loss of an electron:

$$Fe^{++} \longrightarrow Fe^{+++} + e$$

The process of reduction consists of the reverse of these processes, that is loss of oxygen, gain of hydrogen, or gain of an electron. As we shall see later, all these types of chemical change are involved in biological oxidations and each type of reaction is catalysed by a different type of enzyme.

The oxidizing enzymes may be divided into two main groups:

1. Oxidases—catalysing reactions in which molecular oxygen *must* participate.
2. Dehydrogenases which catalyse the transfer of hydrogen.

We must consider in some detail examples of these two groups.

OXIDASES

The oxidases catalyse reactions in which *molecular oxygen* functions as the oxidizing agent or *hydrogen acceptor*. They are of two types, the *aerobic oxidases* which can use oxygen only and which in general form water, and the *aerobic dehydrogenases* which can use, in addition to oxygen, such artificial hydrogen acceptors as methylene blue (MB) and which in general form hydrogen peroxide.

Thus if we represent the substrate to be oxidized by removal of hydrogen as AH_2 which is oxidized to A, the two types of oxidase would act thus:

$$AH_2 + \tfrac{1}{2}O_2 \xrightarrow[\textit{oxidase}]{\textit{aerobic}} A + H_2O$$

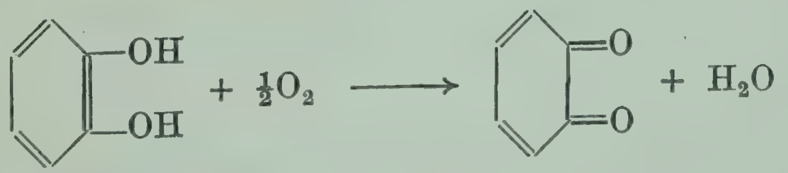
$$AH_2 + O_2 \xrightarrow[\textit{dehydrogenase}]{\textit{aerobic}} A + H_2O_2$$

$$\text{or} \quad AH_2 + MB \xrightarrow[\textit{dehydrogenase}]{\textit{aerobic}} A + MBH_2$$

The **aerobic oxidases** are usually metalloproteins containing iron or copper. They are all inhibited by low concentrations of cyanide which combines with the active group of the enzyme, that is the metal.

I. In the *copper-protein enzymes*, such as phenol oxidases, laccase, and ascorbic acid oxidase, the metal is probably attached directly to the protein.

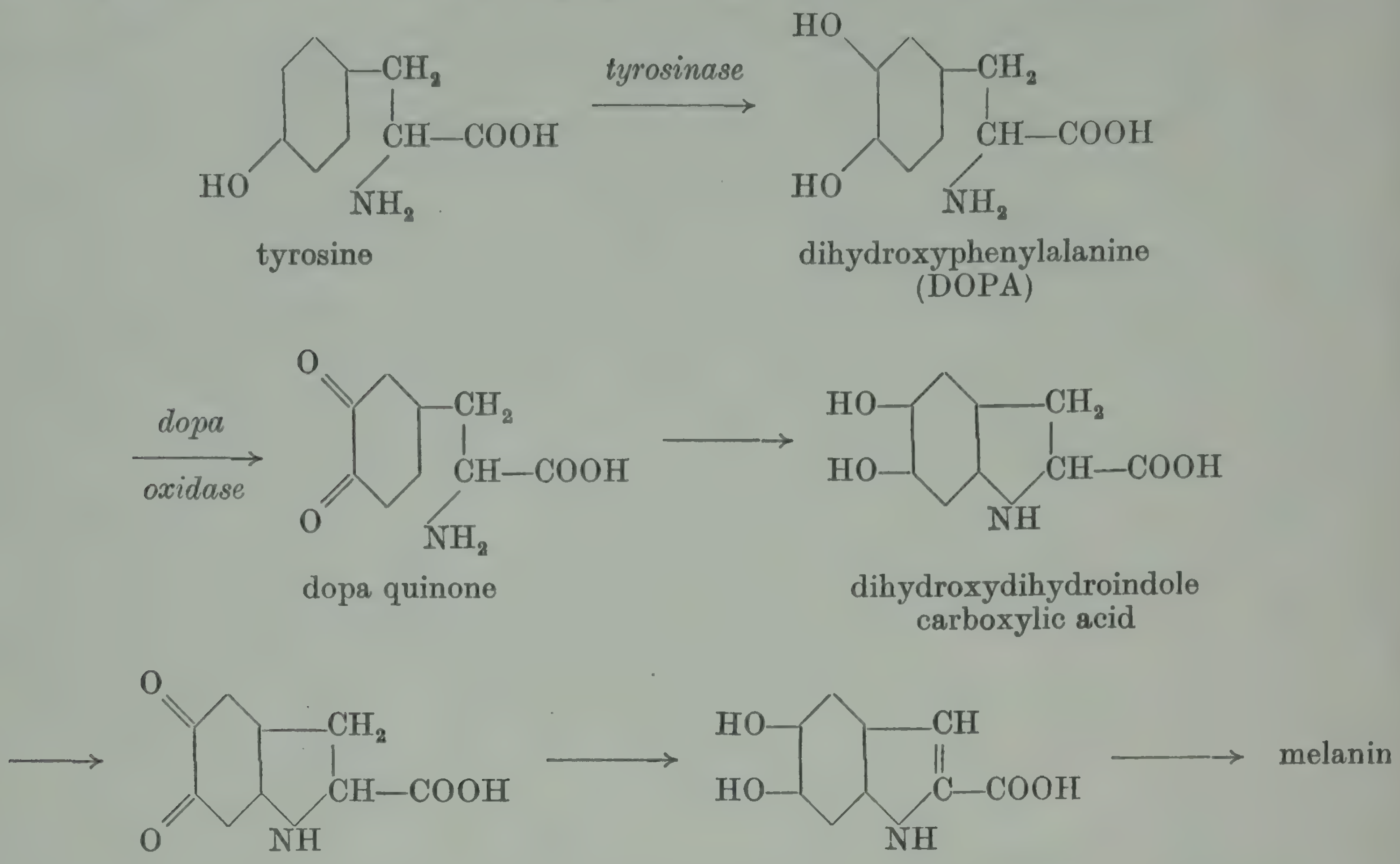
The *phenol oxidase* of the potato, which catalyses the oxidation of such diphenols as catechol, has been highly purified and shown to consist of a protein containing 0·2 per cent. copper. When a purified enzyme preparation is dialysed in collodion sacs against potassium cyanide, the copper combines with the cyanide and can be dialysed away, leaving an inactive protein. Full activity can be restored when this protein is allowed to recombine with copper. It is probable that in the oxidation of catechol to quinone the copper of the enzyme alternates between the cuprous and cupric states, being reduced by the catechol and oxidized by oxygen.



catechol o-quinone

When the cut surface of a potato is exposed to the air for a short time it darkens in colour. This is probably due to the oxidation by the phenol oxidase of diphenols to quinones which then combine with other tissue constituents, for example proteins, to give rise to black pigments.

A very similar phenol oxidase, tyrosinase, has been isolated from a number of plant and animal sources and shown to contain 0·23 to 0·30 per cent. of copper. It oxidizes tyrosine to dihydroxyphenylalanine and this step is the first of a long series of reactions leading to the production of the black pigment melanin (p. 349) which occurs normally in the hair and skin (particularly in the dark races) in the choroid coat of the eye, and pathologically in certain rapidly growing tumours (melanomata).


tyrosinase
tyrosine
dihydroxyphenylalanine (DOPA)
dopa oxidase
dopa quinone
dihydroxydihydroindole carboxylic acid
melanin

The enzyme *ascorbic acid oxidase* which is found in many plants catalyses the oxidation of ascorbic acid (vitamin C) to dehydroascorbic acid (p. 138). It is also a copper protein.

II. In the *iron protein enzymes* the metal is present in the form of an iron-porphyrin complex such as hæmatin (p. 409). The best example is *cytochrome oxidase*, which is discussed below.

The **aerobic dehydrogenases** are not usually metalloproteins: some of them indeed, such as D-amino-acid oxidase and xanthine oxidase, may be flavoproteins (see below). A good example of the aerobic dehydrogenases is to be found in *amine oxidase*, which acts on such amines as tyramine or adrenaline thus:

$$R—CH_2—NH_2 + O_2 + H_2O \xrightarrow[\textit{oxidase}]{\textit{amine}} R—CHO + H_2O_2 + NH_3$$

The hydrogen peroxide which is produced in reactions of this sort is dealt with by the enzymes catalase or peroxidase.

THE DISPOSAL OF HYDROGEN PEROXIDE

Catalase, which is widely distributed in *animal* tissues, is the enzyme which catalyses the decomposition of hydrogen peroxide:

$$H_2O_2 \longrightarrow H_2O + \tfrac{1}{2}O_2$$

As we have just seen, certain steps in biological oxidations involve the formation of hydrogen peroxide which being a strong oxidizing agent would interfere with the delicate oxidation-reduction mechanisms if it were allowed to accumulate. In the presence of catalase, however, hydrogen peroxide is decomposed as fast as it is formed and the catalase therefore exercises a protective action in the cell.

Catalase is an iron protein of molecular weight 248,000 in which the iron is present in the form of hæmatin (p. 409).

Catalase from liver tissue and erythrocytes has been highly purified and obtained in the crystalline state. It is one of the most powerful enzymes known. One molecule of catalase can decompose 2,640,000 molecules of hydrogen peroxide per minute at 0°.

Peroxidase is an enzyme which occurs in nearly all plant cells, but its presence has been reported also in milk and in certain animal tissues, for example spleen. A similar enzyme, *verdoperoxidase*, is found in pus. Highly purified preparations of peroxidase have been split into two components, protein and hæmatin, themselves inactive. When these components are brought together the enzyme is resynthesized and its activity restored.

Peroxidase catalyses the oxidation of a large number of phenols and aromatic amines in the presence of peroxides such as hydrogen peroxide or more complex organic peroxides. For example, in the presence of hydrogen peroxide the juice of the horse-radish, a good source of peroxidase, can oxidize benzidine or guaiacol to blue oxidation products. Peroxidase is a typical enzyme and its activity is destroyed by boiling. Hæmoglobin and its derivatives can also catalyse oxidations by means of hydrogen peroxide even after their solutions have been boiled; they thus behave like thermostable peroxidases. By virtue of their peroxidase-like activity solutions of hæmoglobin derivatives, even after they have been boiled and cooled, can catalyse the oxidation by hydrogen peroxide of benzidine or guaiacol to blue oxidation products. This property is the basis of the well-known guaiacum and benzidine tests for blood; it will be obvious that solutions to be tested for blood must first be boiled to destroy peroxidases.

CYTOCHROME

Cytochrome plays such an important part in oxidation reduction systems in the cell that it must be considered briefly at this stage. This pigment was first discovered by MacMunn in 1886 in muscle and other animal tissues, but little attention was paid to it until it was rediscovered in 1925 by Keilin, who named it cytochrome. It consists of a mixture of three closely related hæmochromogen-like substances, cytochromes *a*, *b*, and *c*, which can be distinguished by means of their absorption spectra. When a strongly illuminated suspension of baker's yeast is examined in the spectroscope the characteristic absorption bands of all three cytochromes are seen. When air is bubbled through the suspension the bands disappear as the cytochromes become oxidized. The bands reappear when the air flow is discontinued and the cytochromes become reduced again.

One of the three components, cytochrome *c*, has been isolated from heart muscle in a highly purified form. It has a molecular weight of 13,000 and contains 0·43 per cent. of iron, the iron being present in the form of hæmatin. The iron can exist both in the ferrous and in the ferric states and it oscillates between these two states as the cytochrome functions in the cell.

Cytochrome *b* can be oxidized spontaneously at a slow rate in the presence of molecular oxygen, but in general the cytochromes are oxidized by molecular oxygen under the influence of the enzyme cytochrome oxidase, formerly known as indophenol oxidase. Cytochrome oxidase is an iron-porphyrin derivative which is thought to be related to cytochrome *a* and is sometimes referred to as a_3. It cannot be obtained in solution as it remains attached to the insoluble particles of the cells when attempts are made to extract it.

Cytochrome oxidase can itself exist in two forms, an oxidized or ferric form (Cyt. oxidase Fe^{+++}) and a reduced or ferrous form (Cyt. oxidase Fe^{++}). The ferrous form is oxidized in the presence of molecular oxygen to yield the ferric form, the oxygen being reduced to water with the aid of two hydrogen ions drawn from the cellular fluids thus:

$$2\ \text{Cyt. oxidase Fe}^{++} + \tfrac{1}{2}\text{O}_2 + 2\text{H}^{+} \longrightarrow 2\ \text{Cyt. oxidase Fe}^{+++} + \text{H}_2\text{O}$$

The ferric form of the cytochrome oxidase is now available to oxidize cytochrome itself.

It is not yet known whether the three cytochromes react separately, or whether they are involved together in a chain of reactions, but it is known that cytochrome *c*, the only soluble member of the group, is the most abundant and therefore presumably the most important. To simplify the discussion which follows it will be assumed that the cytochromes act collectively, being oxidized under the influence of cytochrome oxidase to the ferric form Cyt. Fe^{+++} thus:

$$2\ \text{Cyt. Fe}^{++} + \text{Cyt. oxidase Fe}^{+++} \longrightarrow 2\ \text{Cyt. Fe}^{+++} + \text{Cyt. oxidase Fe}^{++}$$

and being reduced again to the ferrous form Cyt. Fe^{++} by some suitable hydrogen donor, DH_2, thus:

$$\text{DH}_2 + \text{Cyt. Fe}^{+++} \longrightarrow \text{D} + \text{Cyt. Fe}^{++} + 2\text{H}^{+}$$

The nature of the hydrogen donors will be discussed later. The hydrogen ions which appear in this last equation are taken up by the cell fluids.

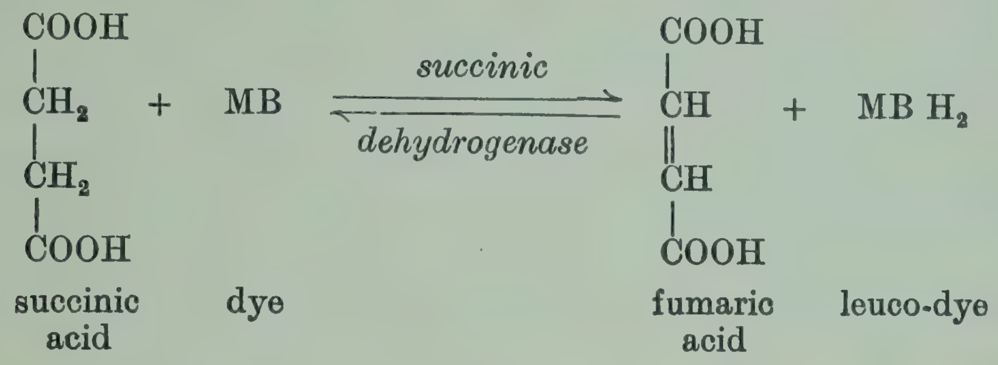
In the presence of small amounts of cyanide (0·001 M) the respiration of many tissues is reduced to about 10 per cent. of its normal value owing to the inhibition of the cytochrome oxidase, and it is probable that in such tissues most of the oxygen utilization goes by way of the cytochrome system. It is possible that the residual respiration after cyanide poisoning involves systems quite different from the cytochromes.

DEHYDROGENASES AND HYDROGEN-TRANSPORTING SYSTEMS

We have already seen that oxidation may involve addition of oxygen, removal of hydrogen, or loss of an electron. The oxygen which is transported by the blood to the cells is used to oxidize the cytochromes. The ferric cytochrome is now left with the task of oxidizing the products of metabolism in the cell which are being broken down to provide energy. In most cases these substances are oxidized by the removal of hydrogen which may be passed from one 'carrier' to another before finally linking up with the ferric cytochrome. The enzymes which catalyse the removal of hydrogen from a substrate are known as *dehydrogenases* or *dehydrases*.

In a few cases the dehydrogenases can connect the substrate directly with cytochrome. An example of such an enzyme is *succinic dehydrogenase*, which oxidizes succinic acid and occurs in muscle and other tissues.

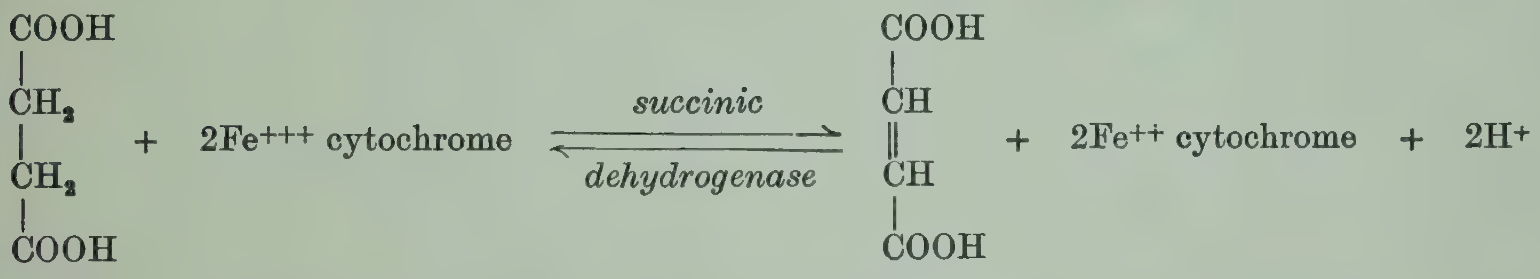
We can demonstrate the function of this enzyme in the laboratory by using the blue dyestuff methylene blue as a 'model' of the hydrogen acceptor systems of the cell. For example, if a small amount of muscle juice containing succinic dehydrogenase is incubated with suitable quantities of succinic acid and methylene blue in a tube from which the air has been evacuated (Thunberg tube), the dyestuff very quickly becomes decolorized by reduction to the 'leuco' form. The enzyme catalyses the transfer of two atoms of hydrogen from the succinic acid to the methylene blue which thus acts as 'hydrogen acceptor':



$$\begin{array}{c}COOH\\|\\CH_2\\|\\CH_2\\|\\COOH\end{array} + MB \underset{dehydrogenase}{\overset{succinic}{\rightleftharpoons}} \begin{array}{c}COOH\\|\\CH\\\|\\CH\\|\\COOH\end{array} + MB\,H_2$$

succinic acid dye fumaric acid leuco-dye

When air is admitted to the tube the atmospheric oxygen oxidizes the reduced dye to the ordinary oxidized blue form. In the living cell the place of methylene blue is taken by ferric cytochrome:

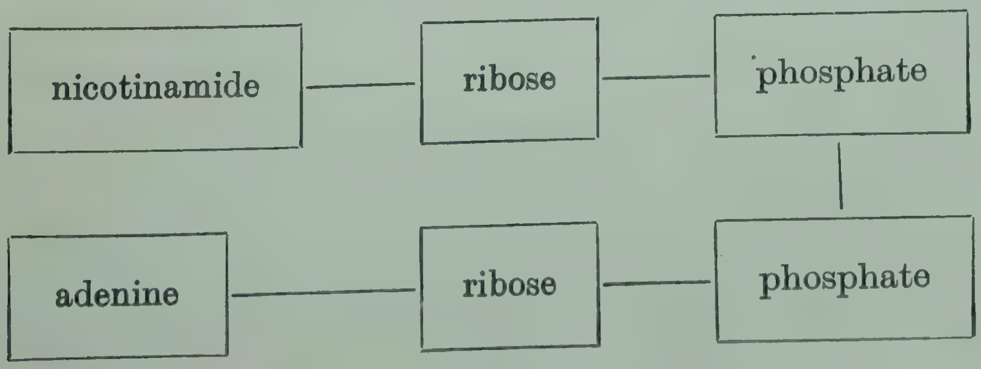


$$\begin{array}{c}COOH\\|\\CH_2\\|\\CH_2\\|\\COOH\end{array} + 2Fe^{+++}\text{ cytochrome} \underset{dehydrogenase}{\overset{succinic}{\rightleftharpoons}} \begin{array}{c}COOH\\|\\CH\\\|\\CH\\|\\COOH\end{array} + 2Fe^{++}\text{ cytochrome} + 2H^+$$

The number of dehydrogenases which can link up a substrate directly with cytochrome in this way is very small. They include succinic dehydrogenase, α-glycerophosphate dehydrogenase and the lactic dehydrogenase of yeast.

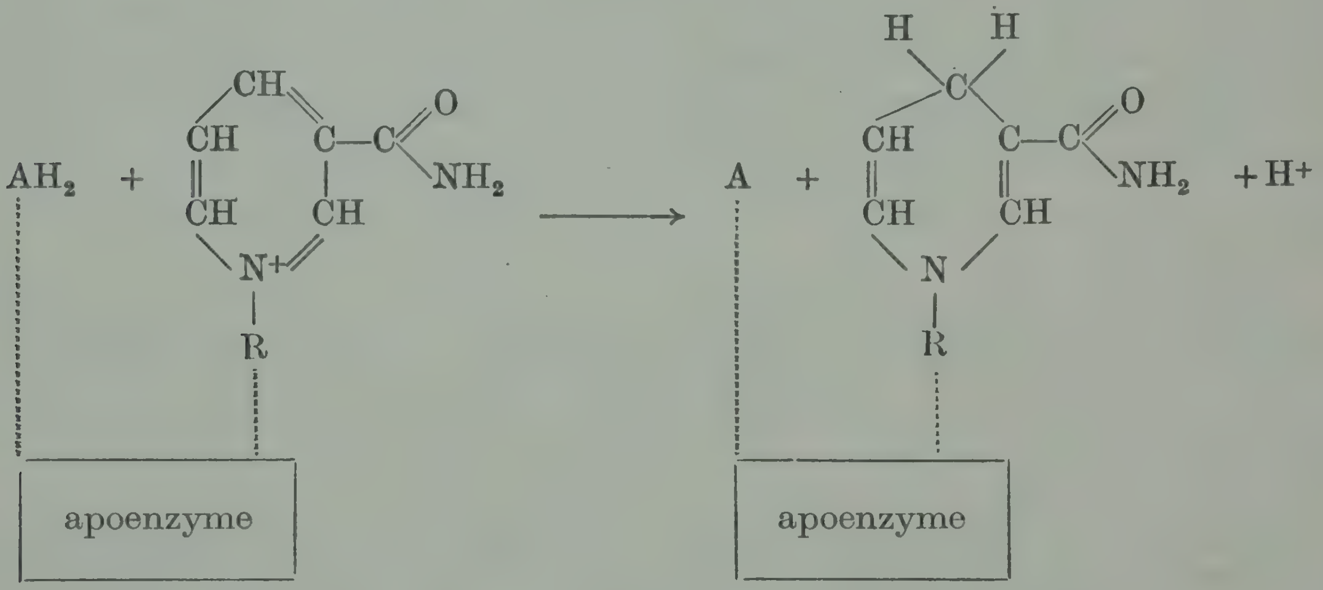
In most cases the hydrogen which is removed from a substrate under the influence of a dehydrogenase is not passed directly to cytochrome but is transferred to a series of hydrogen carriers, which eventually bring it to cytochrome and to oxygen. These carriers include the nicotinamide nucleotides.

The Nicotinamide Nucleotide (pyridine nucleotide) Coenzyme Systems.—The structure of the nucleotides has already been discussed (p. 62). Two nucleotides containing the vitamin nicotinamide (nicotinic acid amide) (p. 127) are known. One of these nicotinamide nucleotides is known as *coenzyme I* (*CoI*), *diphosphopyridine nucleotide* (*DPN*), or *cozymase*. It contains the two bases adenine and nicotinamide and consists of adenylic acid condensed through the phosphate grouping with a nucleotide containing nicotinamide, to form a dinucleotide thus:



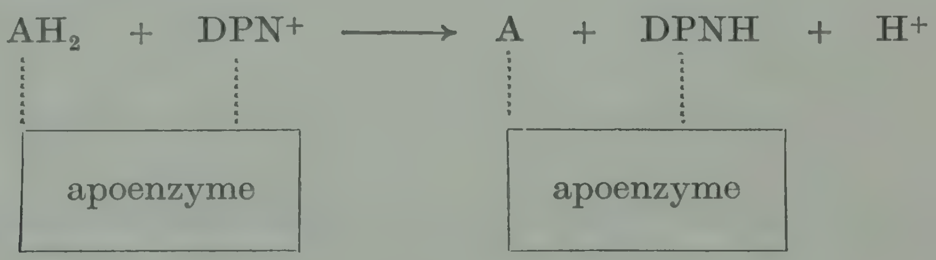
A second such nucleotide is *coenzyme II* (*Co II*), or *triphosphopyridine nucleotide* (*TPN*), which differs from coenzyme I in containing three instead of two phosphoric acid residues in the molecule. These coenzymes exist in the tissues in the free state, but each of them can combine loosely with several different proteins to act as the prosthetic group of the corresponding enzyme system.

Before a substrate AH_2 can be oxidized by one of the coenzyme systems, the presence of a specific protein, the apoenzyme of the dehydrogenase, is essential. This protein forms a fairly loose compound with the nicotinamide nucleotide coenzyme to form a complex which can remove two hydrogen atoms (or more correctly two electrons and two hydrogen ions) from the substrate molecule. Each nicotinamide nucleotide coenzyme can accept hydrogen from a large number of different substrates but for each substrate a specific apoenzyme is essential. Substrate and coenzyme both become attached to the surface of the apoenzyme molecule and the substrate loses two hydrogen ions and two electrons.



R stands for the remainder of the nicotinamide nucleotide molecule.

Or we may simply write:

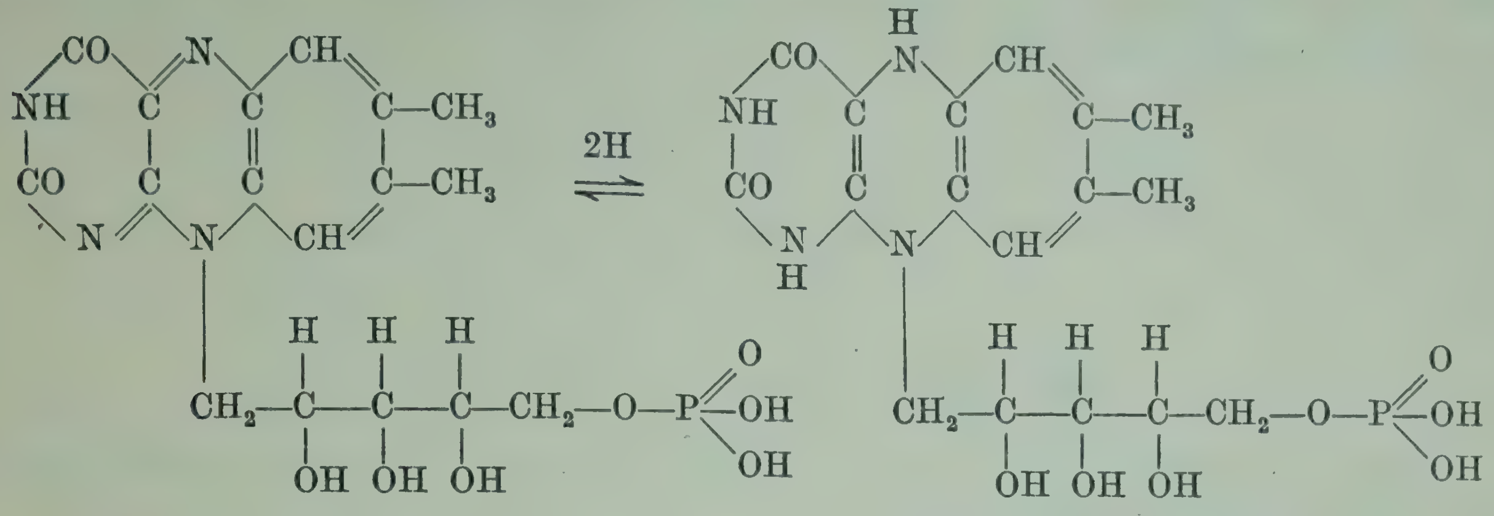


$$AH_2 + DPN^+ \longrightarrow A + DPNH + H^+$$

The two electrons and one hydrogen ion are accepted by the nicotinamide nucleotide; the other hydrogen ion is released into the environment. In this way the substrate AH_2 is oxidized to the oxidation product A while the nicotinamide nucleotide is converted into the reduced form. This reduced form must be reoxidized. Reoxidation is brought about by the *flavoproteins.*

FLAVOPROTEINS

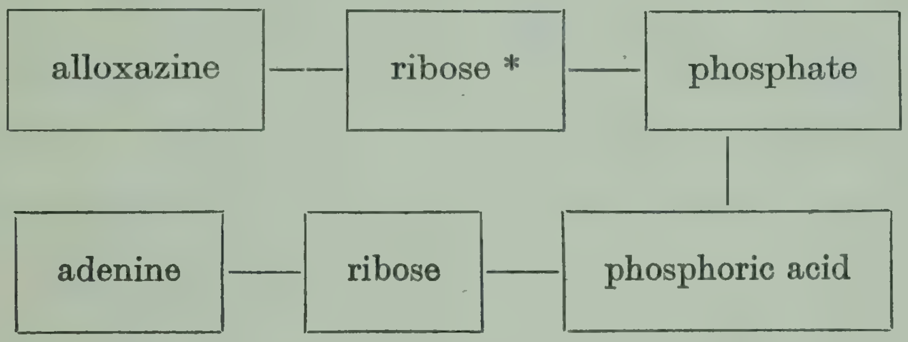
We shall see later that one of the vitamins of the B complex, riboflavin, is converted in the tissues to riboflavin phosphate (alloxazine mononucleotide). This can in turn combine with certain proteins to form the catalytically active *flavoproteins* (*yellow enzymes*) which can exist in both the oxidized and the reduced state as was first shown by Warburg in 1932.



riboflavin phosphate
(alloxazine mononucleotide)

reduced riboflavin phosphate
(dihydroalloxazine mononucleotide)

Riboflavin phosphate condensed with adenylic acid gives the alloxazine adenine dinucleotide (flavin-adenine dinucleotide, FAD), which can also combine with proteins to give flavoproteins. It has already been mentioned in Chapter 6.



alloxazine adenine dinucleotide (FAD).

* Or more correctly ribitol.

Some fifteen flavoproteins are known. Of these three contain alloxazine mononucleotide and seven contain alloxazine adenine dinucleotide. The latter group of flavoproteins includes the D-amino-acid oxidase which catalyses the oxidation of non-naturally occurring optical isomers of the amino-acids, and xanthine oxidase catalysing the oxidation of hypoxanthine to xanthine and of xanthine to uric acid. These enzymes can both be split easily into the protein component and the prosthetic group (alloxazine adenine dinucleotide). Separately the two components are inactive. When they are brought together they recombine and full enzyme activity is restored. The flavoproteins therefore are essentially enzymes in which the flavin is the prosthetic group.

Some flavoproteins contain a metal as well as a flavin pigment. For example, xanthine oxidase contains molybdenum, cytochrome *c* reductase (see below) contains iron, and acyl-coenzyme A-dehydrogenase (see p. 329) contains copper.

In the reoxidation of reduced nicotinamide nucleotides by flavoproteins, two electrons and one hydrogen ion from the reduced pyridine nucleotide are transferred from the environment to the alloxazine ring system of the flavoprotein (FP), converting it to the reduced form (FPH_2):

$$\mathrm{DPNH} + \mathrm{FP} + \mathrm{H}^+ \longrightarrow \mathrm{DPN}^+ + \mathrm{FPH}_2$$

This reduced flavoprotein must now be oxidized by the cytochrome system.

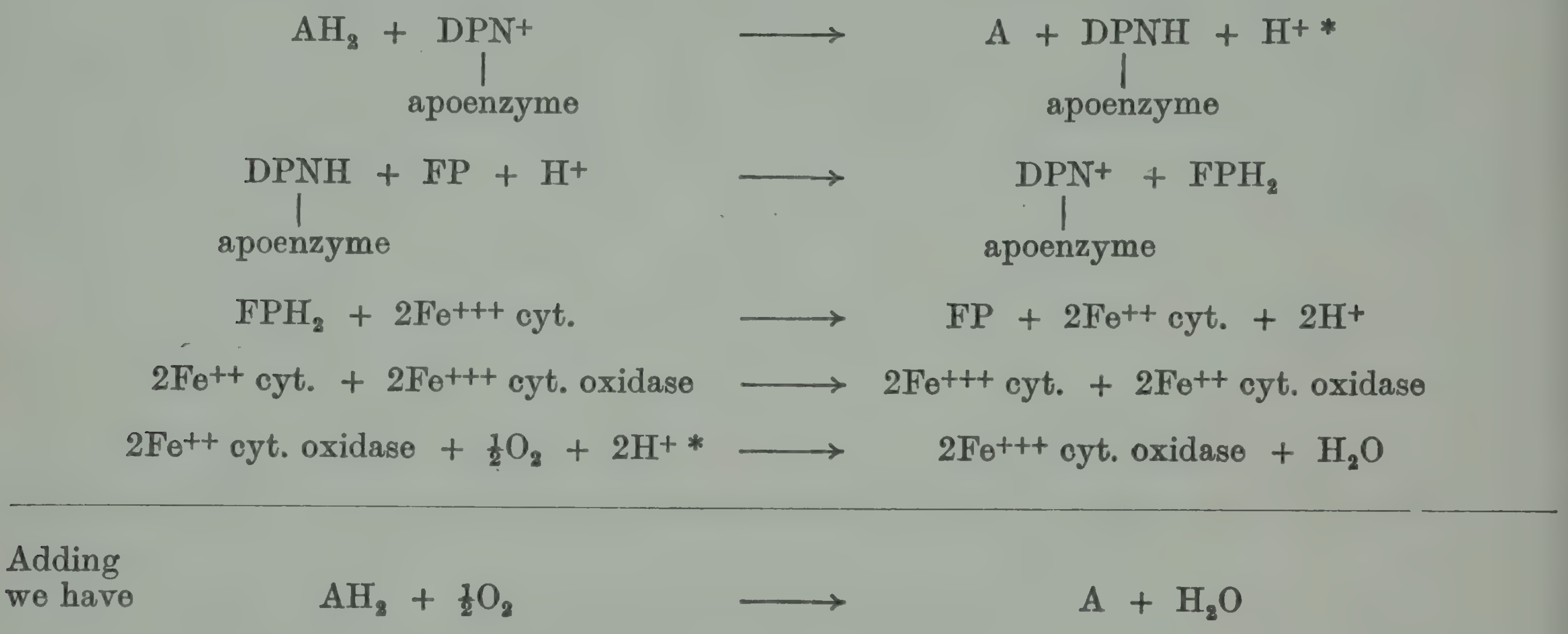
One of the most important flavoproteins is known as *cytochrome c reductase*, which can in its reduced form act as donor of hydrogen to the cytochrome system (that is act as the DH_2 mentioned on p. 84) as shown in the following equation:

$$FPH_2 + 2\ \text{cyt.}\ Fe^{+++} \longrightarrow FP + 2\ \text{cyt.}\ Fe^{++} + 2H^{+}$$

The hydrogen ions are liberated into the cell fluids.

THE MECHANISM OF OXIDATION

The probable mechanism of such a complete system is represented diagrammatically in Fig. 8, 1.



$$\begin{array}{lcl}
AH_2 + \underset{\substack{| \\ \text{apoenzyme}}}{DPN^{+}} & \longrightarrow & A + \underset{\substack{| \\ \text{apoenzyme}}}{DPNH} + H^{+}\,* \\
\underset{\substack{| \\ \text{apoenzyme}}}{DPNH} + FP + H^{+} & \longrightarrow & \underset{\substack{| \\ \text{apoenzyme}}}{DPN^{+}} + FPH_2 \\
FPH_2 + 2Fe^{+++}\ \text{cyt.} & \longrightarrow & FP + 2Fe^{++}\ \text{cyt.} + 2H^{+} \\
2Fe^{++}\ \text{cyt.} + 2Fe^{+++}\ \text{cyt. oxidase} & \longrightarrow & 2Fe^{+++}\ \text{cyt.} + 2Fe^{++}\ \text{cyt. oxidase} \\
2Fe^{++}\ \text{cyt. oxidase} + \tfrac{1}{2}O_2 + 2H^{+}\,* & \longrightarrow & 2Fe^{+++}\ \text{cyt. oxidase} + H_2O \\
\hline
\text{Adding we have}\quad AH_2 + \tfrac{1}{2}O_2 & \longrightarrow & A + H_2O
\end{array}$$

FIG. 8, 1

Schema of oxidation of a substrate AH_2 by a coenzyme system, flavoproteins (FP), cytochrome and molecular oxygen. AH_2 and oxygen are used up and the final products of the reaction are the oxidized substrate A and water. All the other participants go through the cycle of changes repeatedly.

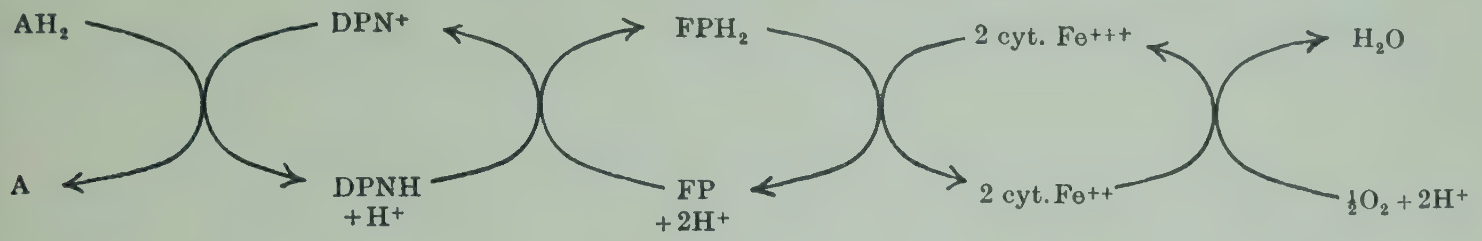
* H^{+} ions pass to and from the environment of the cell.

It will be obvious that a large part of the oxidation process consists of transfer either of hydrogen or of electrons. Hydrogen is transferred from the substrate AH_2 to the nicotinamide nucleotide coenzyme attached to the specific protein A, then on to the flavoprotein and then to cytochrome. Electron transfer occurs among the cytochromes and between them and cytochrome oxidase. Molecular oxygen takes part only at the extreme end of the chain, where it reacts with the cytochrome system.

Such a system as this can bring about the oxidation of a large number of substrates. AH_2 may, for example, be replaced by BH_2 provided that the apoenzyme of the appropriate dehydrogenase is employed.

Moreover, either of the nicotinamide nucleotides (the diphospho- and the triphosphopyridine nucleotides) may function in the system. Diphosphopyridine nucleotide, with a series of appropriate apoenzymes, can bring about the oxidation of such substrates as lactic acid, alcohol, malic acid, hydroxybutyric acid, glucose, formic acid, phosphoglyceraldehyde and glutamic acid. Triphosphopyridine nucleotide, on the other hand, with another set of apoenzymes can bring about the oxidation of 6-phosphoglucose, isocitric acid and glutamic acid.

The sequence of reactions which has just been described may be summarized thus to illustrate the cyclic nature of each stage:


AH_2
A
DPN^+
DPNH $+H^+$
FPH_2
FP $+2H^+$
2 cyt. Fe^{+++}
2 cyt. Fe^{++}
H_2O
$\frac{1}{2}O_2 + 2H^+$

The sequence of events with TPN is exactly analogous. It should be noted, however, that whereas the flavoprotein which connects with TPN is cytochrome reductase, the flavoprotein which connects with DPN is a more complex carrier which requires for its action an additional factor known as the *Slater factor*. The full discussion of this problem is however beyond the scope of this book.

The process of hydrogen transport is summarized in Fig. 8, 2.

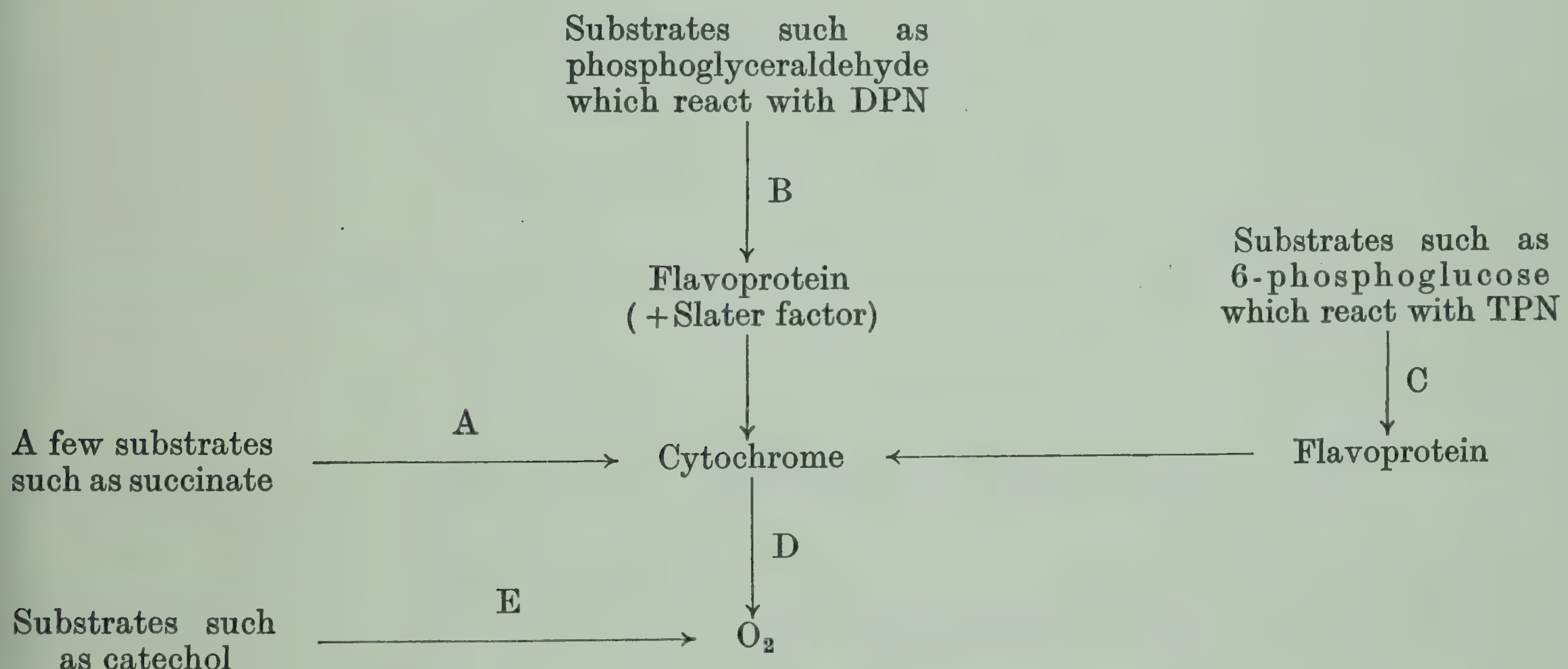

Substrates such as phosphoglyceraldehyde which react with DPN
B
Flavoprotein (+Slater factor)
Substrates such as 6-phosphoglucose which react with TPN
C
A few substrates such as succinate
A
Cytochrome
Flavoprotein
D
Substrates such as catechol
E
O_2

FIG. 8, 2

Main pathways of hydrogen transport. Reaction A is catalysed by such rare enzymes as succinic dehydrogenase, reaction B by dehydrogenases which require DPN as coenzyme, reaction C by dehydrogenases which require TPN as coenzyme, reaction D by cytochrome oxidase, and reaction E by aerobic oxidases such as catechol oxidase.

OTHER HYDROGEN CARRIERS

The tripeptide *glutathione* (glutamylcysteylglycine) may also play a part in hydrogen transport. It is usually written simply as G – SH, where – SH stands for the sulphydryl group of the cysteine and G for the rest of the molecule. In the presence of traces of copper it is readily oxidized by molecular oxygen:

$$2G\text{—}SH + O_2 \longrightarrow G\text{—}S\text{—}S\text{—}G + H_2O_2$$

The oxidized form is readily reduced again in the tissues:

$$G\text{—}S\text{—}S\text{—}G + AH_2 \longrightarrow 2G\text{—}SH + A$$

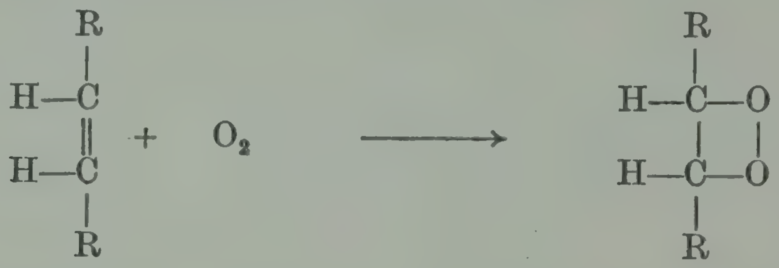
Although glutathione is widely distributed in blood and tissues its use in biological oxidation is not fully understood. It has, however, recently been shown to be the coenzyme of the phosphoglyceraldehyde dehydrogenase described on p. 313, which, of course, also requires DPN. Ascorbic acid (vitamin C) may also act as a hydrogen carrier.

MISCELLANEOUS OXIDIZING ENZYMES

Some oxidizing enzymes cannot easily be classified in any of the previously described groups. Such enzymes include luciferase and lipoxidase.

Luciferase catalyses the oxidation of the substrate 'luciferin' in the presence of oxygen to produce light and oxidized luciferin. This reaction is the basis of the production of light without heat in the glow-worm.

Lipoxidase is responsible for the addition of a molecule of oxygen to the double bonds of certain unsaturated fatty acids, for example linoleic and linolenic acids, to form a peroxide:



The peroxide may then bring about the oxidation of carotenoids and vitamin A.

OXIDATION-REDUCTION POTENTIALS

We have already seen that the process of oxidation involves loss of electrons whereas reduction involves gain of electrons. The reaction in the case of ions may be represented thus:

$$\underset{\text{ferric iron}}{Fe^{+++}} + e \underset{\text{oxidation}}{\overset{\text{reduction}}{\rightleftharpoons}} \underset{\text{ferrous iron}}{Fe^{++}}$$

At equilibrium when the solution contains both ferrous and ferric ions we have in accordance with the law of mass action,

$$\frac{[Fe^{+++}]\,[e]}{[Fe^{++}]} = K,$$

where the brackets denote the concentration of each reactant and K is a constant indicating the oxidation-reduction balance.

If an inert electrode, say of platinum, is immersed in such a solution electrons tending to escape from the ferrous ions give a negative charge to the platinum. If a hydrogen electrode is also immersed in the solution a cell is formed whose potential can be measured. The charge on the inert electrode increases as the tendency of electrons to escape to it becomes greater and is therefore a function of the reducing tendency of the system. The potential in volts set up between the inert electrode and a normal hydrogen electrode is known as the *oxidation-reduction potential*, or *redox-potential*, E_h. It is given by the following equation,

$$E_h = E^\circ + \frac{RT}{nF} \ln \frac{[\text{oxidant}]}{[\text{reductant}]}$$

where E° is the standard e.m.f. of the system.
R is the gas constant.
T is the absolute temperature.
F is the faraday (96,500 coulombs).
n is the number of electron equivalents.
ln is the logarithm to the base e.

When the concentrations of oxidant and reductant are equal $\ln \frac{[\text{oxidant}]}{[\text{reductant}]}$ becomes 0 and $E_h = E^\circ$. When the value of E° is known it is possible to calculate the potential E_h for any degree of oxidation or reduction of the system.

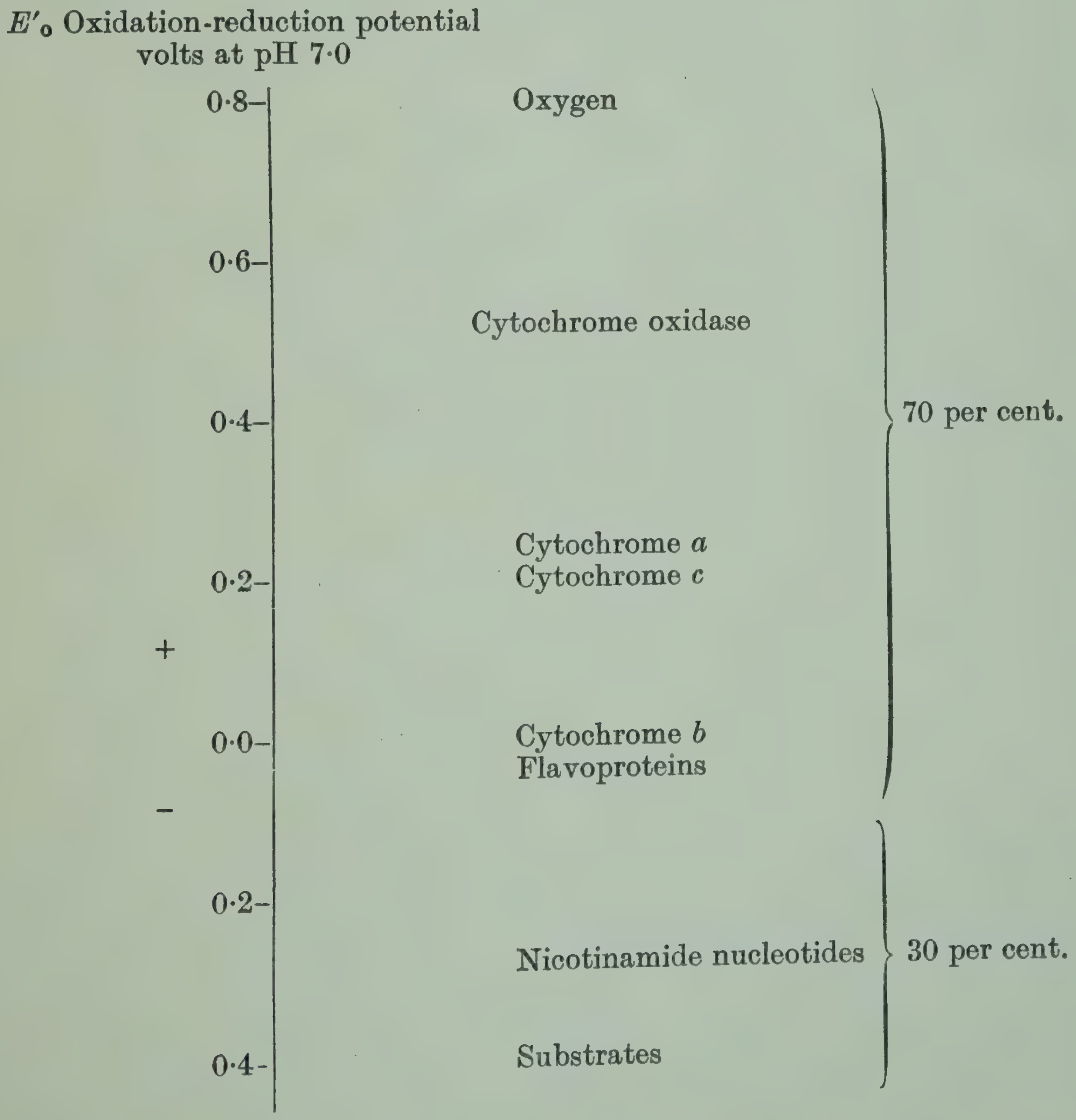

E'₀ Oxidation-reduction potential
volts at pH 7·0
0·8
0·6
0·4
0·2
+
0·0
−
0·2
0·4
Oxygen
Cytochrome oxidase
Cytochrome a
Cytochrome c
Cytochrome b
Flavoproteins
Nicotinamide nucleotides
Substrates
70 per cent.
30 per cent.

FIG. 8, 3

Oxidation-reduction potentials of biological systems. (*After* E. G. Ball.) Each member of the series can oxidize all members below it in the table and is itself oxidized by those above it. Note the correspondence between this figure and Fig. 8, 2. The cytochromes cover 70 per cent. of the potential gap between substrates and oxygen. The hydrogen-transporting systems cover the other 30 per cent.

These relationships apply only when the hydrogen-ion concentration is normal, that is at pH 0. When the pH is at some other value E° becomes E'_0 and the pH value must be quoted.

E'_0 is a measure of the oxidation or reduction *intensity* of a system just as pH is a measure of acid or alkali *intensity*. Thus a system of $E'_0 + 0{\cdot}1$ volt is potentially able to oxidise a system of $E'_0 - 0{\cdot}1$ volt but is itself oxidized by a system of $E'_0 + 0{\cdot}2$ volt.

The E_h value of a system may be measured either by electrical methods, as already mentioned, or by means of indicator dyes (redox indicators). Thus methylene blue, which has an E'_0 value at pH 7·0 of +0·011 volt, is reduced to leucomethylene blue by systems more negative. The dye 2 : 6-dichlorophenolindophenol ($E'_0 = +0{\cdot}217$ at pH 7·0) is used in the quantitative determination of ascorbic acid ($E'_0 = +0{\cdot}06$) which is a powerful reducing agent.

In Fig. 8, 3 is shown the redox-potential range between the substrates in the living cell and the oxygen which is the final oxidant. This gulf is bridged by a series of stepping-stones: the cytochromes, the flavoproteins, the nicotinamide nucleotides. The cytochromes cover 70 per cent. of the potential gap and the hydrogen-transporting systems the remaining 30 per cent.

REFERENCES

BALDWIN, E. (1951). *Dynamic Aspects of Biochemistry*, 2nd ed. Cambridge University Press.

HEWITT, L. F. (1950). *Oxidation-reduction Potentials in Bacteriology and Biochemistry*, 6th ed. Edinburgh: Livingstone.

LARDY, H. A. (Ed.) (1949). *Respiratory Enzymes*. Minneapolis: Burgess Publishing Co.

NEILANDS, J. B. & STUMPF, P. K. (1955). *Outlines of Enzyme Chemistry*. New York: Wiley.

Some aspects of enzyme research (1953). *Brit. med. Bull.* **9**, No. 2.

CHAPTER 9

WATER AND MINERALS

WATER BALANCE

WATER is so well known and so large a constituent of the body that its fundamental importance in both the structure and functioning of all the tissues tends to be overlooked. Deprivation of water brings about death much more rapidly than deprivation of food. When water is completely withheld death occurs in a few days after the body has lost 10 to 20 per cent. of its water content. If, however, water is given but no food, life may continue for several weeks in spite of the loss of all the body fat and 50 per cent. of the tissue protein.

The body of a healthy adult male consists of some 60 to 65 per cent. of water and about 15 per cent. of fat; the remainder is accounted for by the solid parts of cells and supporting structures. There are considerable variations in total body water as between one person and another, which are the result of differences in fat content since the amount of water in the fat-free parts of the body is remarkably constant. This accounts for the fact that the water content of the body of the female is rather less (50 to 55 per cent.) than that of the male. In very fat people the body water may be no more than 40 per cent. of the total body weight. In elderly people and in the fœtus the relative proportion of water in the body is much higher, for example 94 per cent. at the third month of fœtal life. The reasons for these differences are not known.

Recent estimations of the water content of the various tissues of a 70 kg. man are shown in Table 9, 1. Most tissues contain more than 70 per cent. water and even the skeleton and adipose tissues contain quite large amounts. The latter, of course, does not consist entirely of fat; it contains connective tissue, for example, which has water in it and the spaces between the fat cells also contain water. Table 9, 2, also shows that by far the greatest amount of water is to be found in muscle which accounts for the largest part of the body mass.

In its capacity as a solvent water plays a fundamental role in cellular reactions. A very large number of substances are soluble in water and many others, such as fats and fat-soluble compounds, can be carried in fine emulsions or be rendered water-soluble by combination with hydrophilic substances. Certain other properties of water are also of importance. Owing to the high heat capacity of water, large changes in heat production can take place in the body with very little alteration in body temperature. Since the latent heat of evaporation of water is high, the loss of a small amount of water in evaporated sweat means a relatively large loss of heat. Moreover the high latent heat of solidification is a protection against the freezing of the tissues.

In the normal course of events a large amount of water is lost from the body daily and a corresponding amount is taken in, so that water balance is maintained. As shown in Table 9, 3, the amount of water gained and lost by an adult man engaged in a sedentary occupation in a temperate climate is about 2½ litres per day.

TABLE 9, 1

Water Content of Human Tissues

Analytical Figures from one Male Human Body, age 35

Tissue	Per cent. of total Body Wt.	Water Content per cent.
Skin	7·81	64·68
Skeleton	14·84	31·81
Teeth	0·06	5·00
Striated muscle	31·56	79·52
Brain, spinal cord, and nerve trunks	2·52	73·33
Liver	3·41	71·46
Heart	0·69	73·69
Lungs	4·15	83·74
Spleen	0·19	78·69
Kidneys	0·51	79·47
Pancreas	0·16	73·08
Alimentary tract	2·07	79·07
Adipose tissue	13·63	50·09
Remaining tissues—		
Liquid	3·79	93·33
Solid	13·63	70·40
Contents of alimentary tract	0·80	..
Bile	0·15	..
Hair	0·03	..
Total body, weighing 70·55 kg.	100·00	67·85

After H. H. Mitchell, T. S. Hamilton, F. R. Steggerda & H. W. Bean (1945), *J. biol. Chem.* **158**, 625.

The water content of the human red blood cell is 71·7 per cent.

TABLE 9, 2

Percentage of the total Body Water which is found in the various Tissues and Organs

Muscle	50·8	Brain	2·7
Skeleton	12·5	Lungs	2·4
Skin	6·6	Fatty tissue	2·3
Blood	4·7	Kidneys	0·6
Intestine	3·2	Spleen	0·4
Liver	2·8	Rest of body	11·0
			100·0

After E. P. Cathcart (1926), *Glasg. med. J.* **106**, 3.

Water is gained by the body from two main sources. Most of it is taken in by the mouth in the form of food and drink, but a small amount of water is normally formed in the tissues as the result of the oxidation of the hydrogen of foodstuffs. The amount of water ingested in the diet varies, of course, over very wide limits according to habit, climate and occupation. Studies in man with deuterium oxide (heavy water, p. 286) have shown that the absorption of water from the stomach is approximately 2·5 per cent. of the dose per minute whereas from the small intestine it is much more rapid, about 26 per cent. of the dose per minute. Patients with injuries of the limbs absorb water from the alimentary tract more slowly than normal, and in patients with abdominal injuries water absorption is grossly delayed.

In hot climates, where sweating is profuse, the daily loss of fluid may be very high indeed, and a correspondingly high water intake is essential.

TABLE 9, 3

Water Balance

Daily Intake		Daily Output	
Drink	1300 ml.	Urine	1500 ml.
Food	850 ml.	Expired air . . .	400 ml.
Formed in body by oxidation	350 ml.	Skin	500 ml.
		Fæces	100 ml.
Total . .	2500 ml.	Total . .	2500 ml.

Severe muscular activity with associated sweating may also result in considerable water loss and, consequently, increased intake. Playing a hard game of football, for instance, or rowing in a boat race results in the loss of several litres of water in a relatively short time.

The amount of metabolic water formed in the tissues as the result of the oxidation processes described in Chapter 8 is normally about 14 per cent. of the total fluid intake in man, that is about 300 ml. This water is formed in the cells and is of great value to the organism, since its formation is not accompanied by the great osmotic changes associated with the intake of large amounts of fluid. It is of the utmost importance to the hibernating animal which lives for long periods on metabolic water, and to organisms such as the clothes moth which do not normally have access to water.

The water output shown in Table 9, 3, refers to an adult man in a temperate climate. In hot weather the amount of water lost from the skin may be greatly increased, in fact in very hot weather it may be as much as 10 litres, and the volume of urine tends to be reduced (see p. 203). Birds and reptiles conserve water by excreting a small volume of urine consisting of a semi-solid paste of uric acid.

Table 9, 3, gives no idea of the great turnover of fluid which takes place in the body in the course of a day. In twenty-four hours a man secretes 1 to 1·5 l. of saliva; 1 to 2 l. of gastric juice; 0·5 to 1 l. of bile; 0·6 to 0·8 l. of pancreatic juice and 0·2 l. of intestinal juice. All this fluid, with the exception of about 100 ml. which escapes in the fæces, is reabsorbed, so that the fluid output, apart from that lost through the skin and lungs, consists mainly of about 1·5 l. of urine.

The endocrine control of water excretion is discussed in Chapter 35.

MINERALS

The composition of the human body is shown in Table 9, 4. The most abundant elements are of course carbon, oxygen and hydrogen, but minerals such as calcium and phosphorus are also plentiful. Other minerals such as magnesium and iron are present only in small quantities. Nevertheless their presence in the diet is of the greatest nutritional importance. The requirement of a mineral element varies according to age and activity, and is greater in the child, in relation to body weight, than in the adult. It is also increased during pregnancy to meet the needs of fœtal growth, and is still further increased during lactation.

Table 9, 4

Composition of the Human Body (*after Sherman*)

	Wet weight basis per cent.		Wet weight basis per cent.
Oxygen	65	Chlorine	0·15
Carbon	18	Magnesium	0·05
Hydrogen	10	Iron	0·004
Nitrogen	3	Iodine	0·00004
Calcium	1·5	Copper	traces
Phosphorus	1·0	Manganese	
Potassium	0·35	Zinc	
Sulphur	0·25	Fluorine	
Sodium	0·15	Molybdenum, etc.	

The requirement of the adult is determined usually by the *balance technique*. Thus, in the case of calcium, one of the most extensively studied of the minerals, intake may be plotted against output. With increasing intake of calcium, a point is reached at which intake is just adequate to offset excretion, and at this point the subject is in calcium equilibrium. Usually some addition is made to this estimate (for example 50 per cent.) as a safety factor before recommending a minimum safe value to meet human needs. When considering such data it must be understood that the daily intake necessary to balance output varies considerably according to the accustomed intake. In the case of calcium it can, for instance, be increased by giving the subject a diet containing much calcium for some time before the determination is made; conversely, it can be decreased by a diet low in calcium. Furthermore, the daily requirement measured by this method is not necessarily the amount necessary for optimum health.

IRON

The iron content of some common foodstuffs is shown in Table 9, 5. Although in our ordinary mixed diet most of the iron is derived from bread, meat and potatoes, the organ meats such as liver, kidney and heart provide the richest dietary source. Egg-yolk, fish and shell-fish, legumes and green vegetables, nuts, dates, figs, whole wheat, oatmeal and other cereals contain medium amounts. While in some foodstuffs the iron is readily 'available' to the body, in others a high proportion is in organic combination and is not available because it is not absorbed. An appreciable amount of iron may be derived from iron cooking vessels. Milk has a very low iron content. Human milk contains only 0·45 to 1·5 mg. iron per litre and cows' milk usually contains even less. Most mammals are, however, born with a store of iron sufficient to last till weaning. The body of a newly-born child, for example, contains some

TABLE 9, 5*

Iron Content of some Common Foodstuffs. mg. per 100 g. 'edible portion'

Food	Iron mg. per 100 g.
Bread, National, made with 80 per cent. extraction flour	1·2
Oatmeal	4·1
Sirloin	4·0
Heart	8·0
Kidney	13·4
Liver, ox	13·9
Herring	1·5
Milk, fresh, whole	0·1
Butter, Empire, imported	0·1
Cheese, Cheddar type	0·6
Egg, whole, fresh	3·0
Apples, eating	0·3
Currants, black	1·3
Tomatoes	0·4
Prunes	2·9
Lettuce	0·7
Potatoes	0·7
Cabbage	1·0
Carrots	0·6
Chocolate, plain	3·3
Cocoa	14·3

* From *Nutritive Values of Wartime Foods* (1945). *M.R.C. War Mem. No. 14.* London: H.M.S.O.

300 to 400 mg. of iron, largely derived from destruction of its own fœtal hæmoglobin.

During digestion in the stomach the iron of the food is converted into its chloride by the hydrochloric acid of the gastric juice. The iron thus becomes soluble and diffusible in a form which is suitable for absorption from the gut, but before absorption can take place the iron must be reduced to the ferrous state. A part of the iron in the intestinal contents may remain in the ferric state and it may combine with sulphides, phosphates and phytic acid to form insoluble and therefore non-absorbable compounds. Iron in organic combination such as hæmoglobin is not utilized by the body.

The total iron content of the body of an average-sized man is about 4·5 g.; of this 60 per cent. is present in the hæmoglobin of the blood (50 mg. iron per 100 ml.), 20 per cent. is found in myoglobin in muscles and in the iron porphyrin catalysts (catalase, peroxidase and cytochrome), while the remaining 20 per cent. is present in the form of storage iron, which is located chiefly in the liver, spleen and bone marrow. Iron is stored mainly in the form of *ferritin* and *hæmosiderin*. *Ferritin* is a protein which has been isolated as a crystalline derivative from bone marrow, spleen and liver, and from the gastro-intestinal mucosa. This protein may contain as much as 23 per cent., by weight, of iron, the metal being present in the form of the complex $[(FeOOH)_8.FeOPO_3H_2]$, which is bound to protein. Ferritin acts as a store of iron readily available for hæmoglobin synthesis. *Hæmosiderin* is a pigment which has long been known to pathologists in the form of granules or aggregates, visible under the microscope, which can be demonstrated by the blue colour produced when the tissue is immersed in a solution of potassium ferrocyanide (Prussian-blue reaction).

About 1 per cent. of the red blood cells, corresponding to 24 mg. of iron, are destroyed daily, but the iron which is liberated by this breakdown is not excreted: it is stored as ferritin and hæmosiderin and is made use of again in the synthesis of new hæmoglobin. It follows, therefore, that the usual daily requirement of iron is quite small, probably some 1 to 3 mg. in the adult male, and recent experiments with radioactive isotopes have shown that the body absorbs from the gut only sufficient iron to meet its needs. The absorption appears to be regulated by the cells of the gastro-intestinal mucosa according to the needs of the moment, and the ferritin which these cells contain may act as a block controlling absorption (Fig. 9, 6). When the ferrous iron has passed through the mucosa into the blood it is oxidized to the ferric state and becomes attached to serum protein to form *transferrin* which transports iron to the stores in the liver, spleen and bone marrow. This accounts for most of the non-hæmin iron present in blood plasma to an extent of about 120 μg. Fe per 100 ml.

A healthy man on a normal diet absorbs very little iron since his requirement is usually small, and most of the iron in his food passes through the alimentary canal without being absorbed. On the other hand, if the need for iron should increase, as for example after a severe loss of blood, the proportion of dietary iron absorbed from the gut increases considerably. It will be understood from the foregoing that the iron requirements of women are greater than those of men, since the loss of iron at each menstrual period has been estimated to be 4 to 80 mg.

The average mixed diet contains 10 to 50 mg. of iron per day and is probably sufficient to meet most needs. The daily intake recommended by the Committee on Nutrition of the British Medical Association (1950) is 12 mg. per day for adults (Table 12, 17). This amount is probably on the liberal side for men, whose iron requirement is small, but is reasonable for women, whose iron requirement is greater on account of the periodical loss in menstruation. During pregnancy and lactation the iron requirement is increased to at least 15 mg. per day. If these requirements are not met anæmia will ultimately result. Such iron-deficiency anæmia is quite common among women, especially in the lower-income groups, and can be cured by the administration of iron salts such as ferrous sulphate.

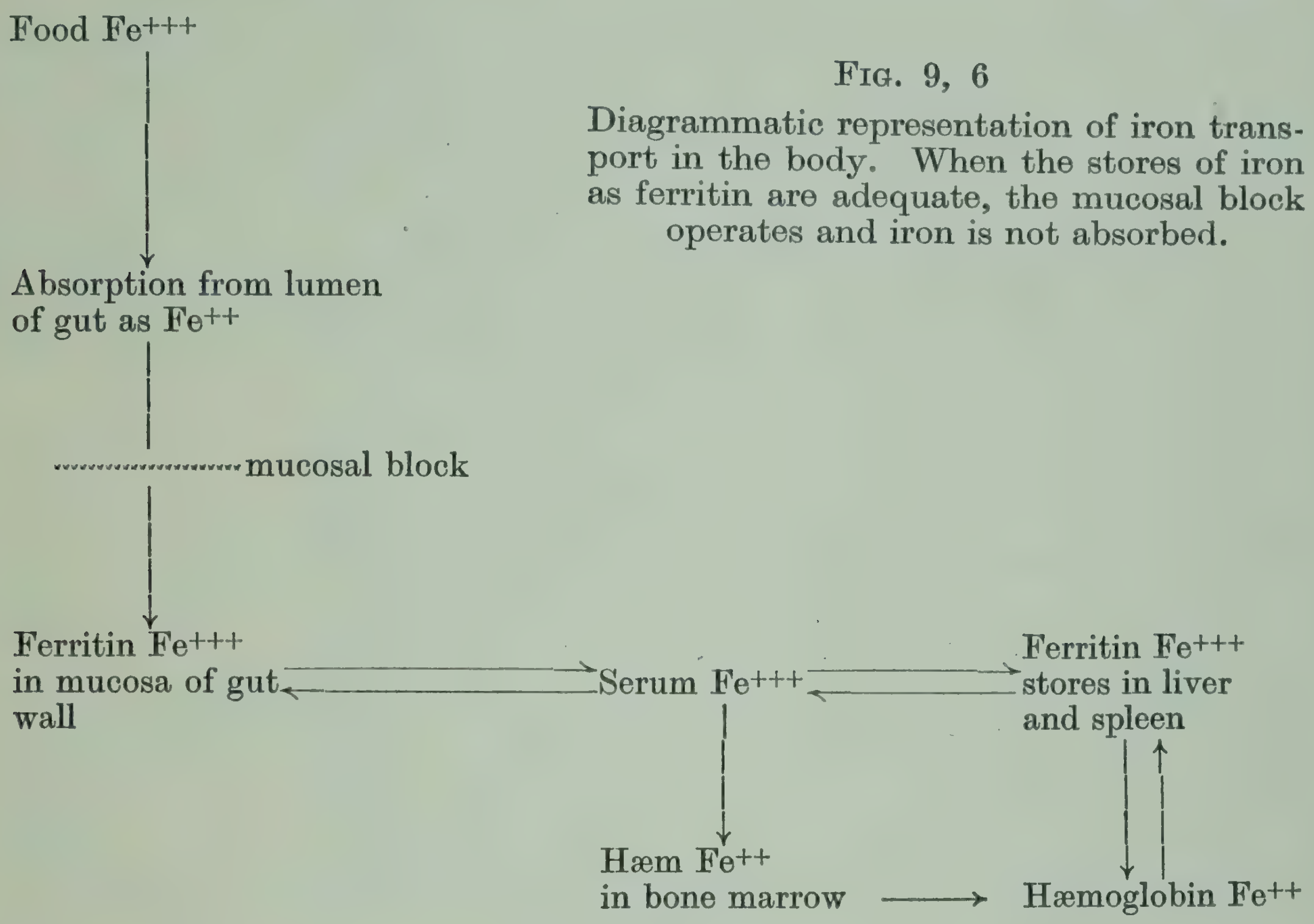

Food Fe^{+++}
Absorption from lumen of gut as Fe^{++}
mucosal block
Ferritin Fe^{+++} in mucosa of gut wall
Serum Fe^{+++}
Ferritin Fe^{+++} stores in liver and spleen
Hæm Fe^{++} in bone marrow
Hæmoglobin Fe^{++}

FIG. 9, 6

Diagrammatic representation of iron transport in the body. When the stores of iron as ferritin are adequate, the mucosal block operates and iron is not absorbed.

The normal adult excretes 1 to 4 mg. of iron per day in the urine. The bulk of the iron excreted is found in the fæces and it was originally believed that iron was excreted into the bowel. This view is no longer held, since it appears that nearly all the iron of the fæces represents iron unabsorbed from the diet. Significant amounts of iron may be lost in the sweat and in the hair.

It has been established that in some species of animal iron alone is insufficient to ensure an adequate supply of hæmoglobin for which minute amounts of copper are also necessary.

CALCIUM

The calcium content of some common foodstuffs is shown in Table 9, 7. Calcium is ingested mainly in milk, cheese and vegetables; but other good sources are eggs, butter, oranges, nuts and carrots. An appreciable amount of calcium is obtained from the drinking water in some districts. The average figure for Britain as a whole for the calcium obtained from the water supply is 75 mg. per person per day, but the intake is very much greater than this in districts where the water is hard.

Calcium is absorbed with difficulty from the gut and in general only about half of the ingested calcium is absorbed. The absorption of calcium depends upon the acidity of the intestinal contents. In a slightly acid medium calcium forms soluble salts which can be absorbed, but in neutral or alkaline solution calcium tends to be precipitated as an insoluble phosphate or phytate. Calcium absorption, therefore, is mainly confined to those regions of the intestine where the contents are slightly acid (p. 254).

TABLE 9, 7

Calcium and Phosphorus Content of Foods *

Foodstuff	Calcium mg. per 100 g.	Phosphorus mg. per 100 g.
Bread, white	23·1	73
Bread, wholemeal	30·0	213
Oatmeal, raw	55·3	380
Butter, fresh	14·8	24
Cheese, Cheddar	810·0	545
Eggs, raw or boiled	56·0	218
Milk, fresh whole	120·0	95
Beef, sirloin, roast, lean and fat	6·5	284
Cod, steamed	14·6	242
Herring, fried	38·6	339
Sardines, tinned	409·0	683
Apples, Empire, eating	3·6	6·8
Currants, black, raw	60·3	43·2
Oranges	41·3	23·7
Rhubarb, raw	103·0	21·0
Almonds	247	442·0
Beans, butter, raw	84·8	318·0
Beans, haricot, raw	180·0	309·0
Carrots, young, boiled	28·8	29·5
Peas, fresh, boiled	12·6	83·3
Potatoes, old, boiled	4·3	29·0
Turnips, boiled	55·0	19·2

* From R. A. McCance & E. M. Widdowson (1940), Chemical composition of foods. *Spec. Rep. Ser. med. Res. Coun., Lond.* No. 235. London: H.M.S.O.

The absorption of calcium is defective when the vitamin D intake is low, and in rickets (p. 144) calcium absorption is negligible.

In the blood, calcium is found in the plasma only, none is present in the cells. Normal human blood plasma contains 9·5 to 11·5 mg. of calcium per 100 ml. (Table 9, 8). Of this total quantity about 55 per cent. is present in the form of calcium ions, while the remainder is present in the form of non-diffusible complexes of calcium with protein. The calcium concentration in the cerebrospinal fluid is 5 to 7 mg. per 100 ml. The relationship of the diffusible to the non-diffusible calcium in the blood depends upon the concentration of plasma proteins; consequently in diseases in which the plasma protein level is low, for example chronic kidney disease, the total amount of calcium in the plasma may fall to low levels. Tetany (Chapter 51), however, does not occur because in such cases the fall in calcium is confined to the non-ionized fraction.

The concentration of calcium in the blood plasma is maintained within rather narrow limits by a number of regulating factors. For example, it depends to some extent upon the amount of calcium absorbed from the gut, and further upon the vitamin D content of the diet. Thus in cases of rickets, where the calcium absorption is defective, the calcium level of the plasma or serum may fall as low as 7 mg. per 100 ml., and conversely in patients who have received very large doses of vitamin D (hypervitaminosis D) the level may reach 15 mg. per 100 ml.

The blood calcium level is also controlled by the parathyroid glands in a way which is not yet fully understood. If the parathyroid glands are removed, as sometimes happens accidentally in surgical operations for the removal of the thyroid, the calcium level of the blood falls, and if it reaches a sufficiently low level the condition of tetany may result (Chapter 51). On the other hand, a tumour of the parathyroid may produce excessive amounts of the hormone and this is followed by a rise in the blood calcium concentration. This increase is not due to increased absorption of calcium, or to diminished excretion, but results from the mobilization of calcium from the bones, which become weak, easily fractured and deformed (*generalized osteitis fibrosa*). The concentration of calcium in the blood plasma may rise to some 15 or 16 mg. per 100 ml., large amounts of calcium are excreted into the urine, and calcium salts may be deposited in the kidneys and in the walls of the larger blood vessels (see also Chapter 51).

TABLE 9, 8

Calcium Concentration in Normal Human Blood Serum

	mg. per 100 ml.
Total serum calcium	9·5 to 11·5
Non-diffusible (bound to serum protein, mainly albumin)	4·0 to 5·0
Diffusible	5·6 to 6·5
Ionized	4·8 to 6·3

After A. McLean & A. B. Hastings (1935), *Amer. J. med. Sci.* **189**, 601.

Calcium is excreted mainly by the bowel, partly by the distal ileum and to a slight extent in the colon, so that the fæces contain not only that calcium of the food which has not been absorbed but also a certain amount which has been re-excreted into the gut. Between 100 and 300 mg. of calcium are excreted daily in the urine.

The main function of calcium in the body is as a structural material in the bones and the teeth (p. 104). If the calcium intake is inadequate over a long period the bones are reduced in thickness, and since they are not so strong as normal bones they break more readily. Since the skeleton acts as a reservoir from which calcium may be drawn to provide for the various needs of the body, the bones may become thin during pregnancy and lactation in cases where the calcium intake is too low to supply the needs of the growing fœtus or to compensate for the loss of calcium in the milk.

Calcium is an important factor in the regulation of the excitability of the nerves. As is shown in Chapter 51, a low blood calcium level is associated with hyperexcitability of the nervous system (tetany). Calcium is also essential for the normal action of heart and skeletal muscle.

Calcium is required in the process of clotting of blood (p. 361) and of milk (p. 174).

The calcium requirements in the diet are relatively high and are of the order of one gram per day. The British Medical Association recommendations (Table 12, 17, p. 190) give a figure of 0·8 g. daily for adults and of 1 g. daily for young children. For growing children of 11 to 19 the requirement varies between 1·1 and 1·4 g. During pregnancy and lactation the requirement is increased to 2 g. per day to provide for the growth of the child *in utero* and later to provide the calcium of the milk (about 1 g. calcium per litre).

PHOSPHORUS

The phosphorus content of some common foodstuffs is shown in Table 9, 7. Phosphorus is ingested in the form of phosphate, both inorganic and organic, the latter including phosphoproteins (for example caseinogen of milk), phosphatides and nucleic acids. The most important sources are milk, cheese, eggs, meat and fish, but it is also fairly abundant in cereals and in certain vegetables. A considerable but variable proportion of the phosphorus in cereals is present in the form of phytic acid (inositol hexaphosphate, p. 13). This substance is not hydrolysed in the gut and is not diffusible. Its phosphorus is, therefore, not available to the body and is excreted unchanged. Hence in the calculation of the available phosphorus of any foodstuff phytic acid phosphorus should be allowed for. The phytic acid content of some common foodstuffs is shown in Table 9, 9.

TABLE 9, 9

Phytic Acid Content of Foods *

	mg./100 g.
Oatmeal	264
Wheatmeal, 92 per cent. extraction	214
Wheatmeal, 85 per cent. extraction	127
White flour, 70 per cent. extraction	51
Wholemeal bread (92 per cent.)	88
'National Wheatmeal' bread (85 per cent.)	42
White bread (72 per cent.)	11
Harvest beans, butter beans, dried peas (raw)	229
Almonds, Brazil nuts	176

* Modified from *Manual of Nutrition* (1945), Ministry of Food, London: H.M.S.O., and E. M. Widdowson (1941), *Nature,* **148,** 219.

Phytic acid not only accounts for a loss of part of the dietary phosphorus, but because it forms insoluble salts with calcium and with iron it tends to interfere with calcium and iron absorption. It is to allow for such losses that calcium carbonate is now added to flour used in the preparation of National bread (p. 180). Wheat, but not oats, contains the enzyme phytase which catalyses the hydrolysis of phytic acid to inositol and phosphoric acid. In bread-making a considerable proportion of the phytic acid in flour may be broken down by the enzyme while the dough is standing, and the phosphorus is thus rendered available.

Phosphorus in the form of inorganic phosphate is absorbed readily from the gut, but in an alkaline medium it tends to be precipitated as insoluble phosphates such as calcium phosphate. Phosphorus absorption, therefore, occurs principally in those regions of the intestine where the reaction is slightly acid.

Table 9, 10

Phosphorus Compounds in the Blood of Men and Women fasted overnight *

	mg. per 100 ml. blood (as phosphorus)
Inorganic phosphate . .	2·5 to 3·8
Organic or ester phosphate .	20·3 to 25·6
Lipid phosphorus . . .	10·4 to 13·1

* From H. D. Kay & J. Byron (1927), *Brit. J. exp. Path.* **8**, 240.

The concentration of the various compounds of phosphorus in normal human blood is shown in Table 9, 10. The inorganic phosphate present in about the same concentration in whole blood and plasma is the fraction of the most clinical significance. Both the lipid phosphorus and the organic or ether phosphate are more abundant in the red cells than in the serum.

Phosphorus is excreted mainly by the kidneys and about two-thirds of the ingested phosphorus appears in the urine (1 to 5 g. P_2O_5 per day). The remainder is excreted in the fæces. The elimination of inorganic phosphate in the urine is impaired in chronic disease of the kidneys which therefore may be accompanied by an increased concentration of phosphate in the serum.

An important function of phosphorus, like that of calcium, lies in the formation of the bones and the teeth, and its role in calcification is discussed on p. 105. It is also required for the formation of the phosphatides and the nucleic acids which are essential structural components of the cell, and for the formation of organic esters which play an important part in the breakdown of carbohydrate in the tissues (p. 294). It is utilized to some extent in the absorption of monosaccharides from the gut, and the phosphates of the blood and tissue fluids play an important role as buffers. The phosphorus requirement in the diet is of the order of 1 to 2 g. per day; more is needed during pregnancy and lactation.

Not only the absolute amount of calcium and phosphorus, but also the ratio of calcium to phosphorus, in the diet is of some importance. The National Research Council of the United States recommends an optimum ratio of calcium to phosphorus of 1 : 1·5 for adults and 1 : 1·0 for children and for pregnant and lactating women. In general a diet supplying sufficient calcium also supplies an appropriate amount of phosphorus.

The Phosphatases

Several different types of phosphatase are known, but the commonest is the phosphomonoesterase which breaks down simple organic phosphates such as phosphoglucose and phosphoglycerol with the liberation of free inorganic phosphate.

$$\text{6-Phosphoglucose} \xrightarrow{\textit{phosphatase}} \text{glucose} + \text{phosphoric acid}$$

$$\text{Phosphoglycerol} \xrightarrow{\textit{phosphatase}} \text{glycerol} + \text{phosphoric acid}$$

Alkaline phosphatase, most active about pH 9, occurs in many tissues but especially in bone, calcifying cartilage and in the kidney. Acid phosphatase, most active about pH 4, is found in the prostate gland and in very high concentration in seminal fluid (87,000 to 330,000 units per 100 ml.). Red blood cells contain a third phosphatase, most active about pH 6. The last is, however, active over a wide pH range, and it is therefore important that specimens of blood used for determination of acid or alkaline phosphatase should be quite free from hæmolysis.

Both acid and alkaline phosphatase are present in the blood plasma and their concentration is usually expressed in terms of King-Armstrong phosphatase units. Normal blood plasma contains 1 to 5 units of acid phosphatase and 3 to 10 units of alkaline phosphatase per 100 ml., but these values may be considerably altered in disease. Thus in obstructive jaundice the concentration of alkaline phosphatase may rise to over 100 units per 100 ml., while that of acid phosphatase remains unchanged. An elevated concentration of alkaline phosphatase in the blood constantly accompanies increased bone formation. On the other hand, in cases of cancer of the prostate gland, in which secondary tumour deposits are found in the bones, a high concentration of acid phosphatase is found in the blood plasma.

The Composition of Bones and Teeth

Calcium and phosphorus are important constituents of bones and teeth. Bone is composed partly of inorganic material (60 per cent.) and partly of an organic matrix (40 per cent.) consisting chiefly of the protein *ossein*, which resembles the collagen of cartilage or tendon. Bone contains, in addition to calcium and phosphorus, small amounts of magnesium, sodium, potassium, strontium and fluorine. It is usually considered that the mineral in bone exists as an apatite, a double salt of calcium phosphate and calcium carbonate. Citric acid is abundant in bone but its significance is not understood.

Teeth have a similar composition but their water content is lower than that of bone. Enamel contains only 5 per cent. of water, dentine 10 per cent., while bone contains about 25 per cent. of water. Both dentine and enamel contain a higher concentration of calcium phosphate and less calcium carbonate than bone.

Contrary to earlier opinions, the mineral salts of bone are now known to be metabolically active. This can be demonstrated with the use of radioactive isotopes. For example, Hevesy fed rats small amounts of sodium phosphate containing radioactive P as indicator. By means of tissue analysis and mineral balance studies he found that the average time a P atom spent in the body was about two months. Even in the skeleton of the adult rat about 30 per cent. of the P atoms deposited were removed in the course of twenty days. Since in adult rats there is virtually no further growth of bone, the experiment indicates that the mineral matter of bone is continually changing. Even the molar teeth, which do not grow in adult rats, took up some of the radioactive P.

CALCIFICATION

During the process of calcification bone salts are laid down in the calcifying matrix, but the reason for the deposition of bone salts at

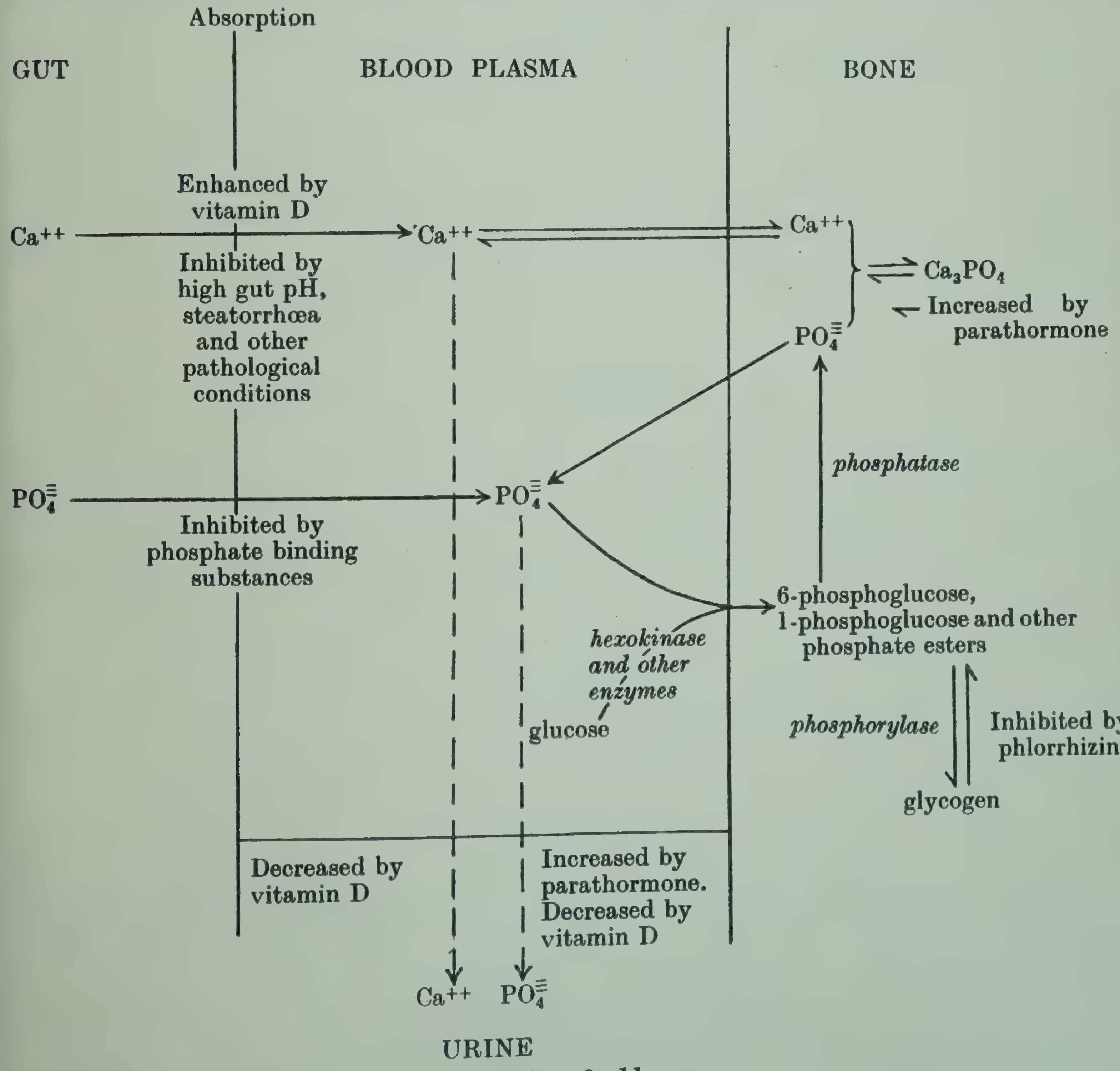

Absorption
GUT
BLOOD PLASMA
BONE
Enhanced by vitamin D
Ca++
Ca++
Ca++
Ca3PO4
Increased by parathormone
Inhibited by high gut pH, steatorrhœa and other pathological conditions
PO4≡
phosphatase
PO4≡
PO4≡
Inhibited by phosphate binding substances
6-phosphoglucose, 1-phosphoglucose and other phosphate esters
hexokinase and other enzymes
glucose
phosphorylase
Inhibited by phlorrhizin
glycogen
Decreased by vitamin D
Increased by parathormone. Decreased by vitamin D
Ca++
PO4≡
URINE

FIG. 9, 11

Schematic representation of calcium and phosphorus metabolism.

certain sites only and not generally throughout the body is not completely understood. In order to explain this phenomenon Robison carried out some important experiments with calcifying cartilage. He found that if a young growing bone was split and immersed in a solution of calcium chloride, or of sodium phosphate, no calcification occurred, but if the bone was immersed in a solution containing calcium glycerophosphate, or calcium hexose phosphate, calcification proceeded normally at the ossifying centres. He assumed therefore that the enzyme phosphatase, which is abundant in the cells of the ossifying zone, hydrolysed these organic esters with the production of a high local concentration of inorganic phosphate and the consequent deposition of insoluble calcium phosphate.

It has been suggested that blood plasma is saturated with respect to calcium phosphate, and therefore any increase either in the calcium-ion concentration or in the phosphate-ion concentration tends to cause precipitation of insoluble calcium phosphate. Phosphate-ion concentration is presumed to be increased in tissues in which calcification is active owing to the action of the enzyme phosphatase in hydrolysing phosphate esters which are particularly abundant in such tissues. In calcifying cartilage, for example, it has been postulated that as the result of the process of glycolysis (p. 311) glycogen is acted upon by the enzyme phosphorylase in the presence of plasma inorganic phosphate to form 1-phosphoglucose and related esters such as 6-phosphoglucose which are then hydrolysed by bone phosphatase with the localized liberation of high concentrations of phosphate ions (Fig. 9, 11). These, in turn, react with calcium ions with the deposition of bone salts. An absence of a similar high concentration of phosphate esters accounts for the fact that in other tissues, such as the intestine and kidney, where the local concentration of phosphatase is also high, precipitation of insoluble calcium salts does not occur.

MAGNESIUM

The importance of magnesium in the diet is less obvious than that of calcium since the magnesium intake is seldom deficient. A normal diet contains 0·4 g. per day. Like potassium, magnesium is found mainly inside cells; thus the concentration of magnesium in the intracellular fluid is about 28 m.equiv per litre while in the extracellular fluid (plasma) its concentration is from 1·6 to 3·2 m.equiv per litre (2 to 4 mg. per 100 ml.). About three-quarters of the magnesium present in the body is found in the bones. Magnesium ions are essential for the functioning of certain enzymes, for example phosphoglucomutase (p. 295).

IODINE

Iodine is an essential element in the diet since it is required for the synthesis of the hormone of the thyroid gland. The chief sources of iodine are inorganic iodides in drinking water, and in such foods as vegetables and fish. In certain parts of the world where the soil and, consequently, the water supply and vegetation are poor in iodine a condition known as *simple colloid goitre* may occur, in which the thyroid gland is greatly enlarged (Chapter 51). This condition was endemic in certain parts of the world, usually remote from the sea, such as Derbyshire, in certain cantons in Switzerland, and in the area of the Great Lakes in North America, but it has now been largely prevented by adding minute amounts of iodide to the drinking water or to the table salt. The iodine requirement of a normal adult is about 0·15 mg. per day. Children require more than this amount, and increased requirements are also found at puberty, during pregnancy and lactation, and at the menopause.

FLUORINE

Fluorine is present in animal soft tissues in amounts ranging from 0·2 to 1·5 mg. per 100 g. of the dry material. Bones and teeth contain larger amounts (20 to 30 mg. per 100 g. ash) and the dentine of teeth may contain much more than the enamel. The incidence of caries tends to be lower where the drinking water contains a small amount of fluorine but in districts where the water supply contains unusually large amounts of fluorine the enamel of the teeth may show a characteristic mottling and the bones may be greatly thickened.

SODIUM

The total amount of sodium in the body of a man of 70 kg. is about 2800 m.equiv, or 65 g. The body's main source of sodium is the sodium chloride in the diet. There is great variation in the average daily intake of salt, from 10 to 30 g. per day, for some persons habitually take much more salt than others. If salty foods are avoided and no salt is added to the food during cooking or at the table the daily intake of salt can be reduced to about 4 g., but further reduction requires the preparation of salt-free bread and other special measures. Sodium is readily absorbed and distributed mainly throughout the extracellular fluid, though some is present in the cells and also in bone. The concentration of sodium ion in normal plasma is 137 to 148 m.equiv per litre (318 to 342 mg. per 100 ml.).

The sweat contains some 50 to 100 m.equiv of sodium per litre and there is a negligible amount lost in the fæces. In a temperate climate, therefore, the only important loss is normally through the kidney; this is dealt with more fully in Chapter 35.

POTASSIUM

In contrast to sodium, which is found mainly in the extracellular fluid, potassium is principally inside the cells (about 150 m.equiv per litre). The concentration of potassium ion in the plasma is 3·9 to 5·0 m.equiv per litre (15 to 20 mg. per 100 ml.), but a much higher concentration is found within the red cells. A practical implication of this fact is that even a trace of hæmolysis can falsify determinations of the serum potassium.

TRACE ELEMENTS

Minute amounts of metals such as copper, cobalt, molybdenum and manganese are essential components of the diet for normal human and animal nutrition and their absence may result in deficiency disease. For example cattle and sheep develop 'pine' if cobalt is missing from the pasture while newborn lambs suffer from 'swayback' in the absence of copper. Some of these metals, for example copper and zinc, are constituents of certain enzymes, while cobalt is a constituent of vitamin B_{12}.

REFERENCES

GABLER, C. J. (1956). Absorption and metabolism of iron. *Science*, **123**, 87–90.

GRANICK, S. (1951). Structure and physiological functions of ferritin. *Physiol. Rev.* **31**, 489–491.

GRANICK S. (1954). *Bull. N.Y. Acad. Med.* **30**, 81–105.

HAHN, P. E. (1948). The metabolism of iron. *Fed. Proc.* **7**, 493–498.

McCOLLUM, E. V. (1941). *Inorganic Elements in Nutrition.* University of Pennsylvania Press.

NICOLAYSEN, R. & EEG-LARSEN, N. (1953). The biochemistry and physiology of vitamin D. *Vitamins and Hormones*, **11**, 29–41.

NICOLAYSEN, R., EEG-LARSEN, N. & MALM, O. J. (1953). Physiology of calcium metabolism. *Physiol. Rev.* **33**, 424–444.

PIRRIE, R. (1950). Iron metabolism. *Glasg. med. J.* **31**, 397–412.

CHAPTER 10

THE VITAMINS

As our knowledge of the metabolic role of the vitamins grows it becomes increasingly difficult to define what is meant by a vitamin in a formula which is both concise and complete. At present perhaps the best definition is the old schoolboy one: "vitamins are things which if you don't get in small daily doses in your food make you ill." Although open to criticism on scientific as well as on syntactical grounds such a definition at least has the merit of emphasizing that the end-result of vitamin deficiency is disease.

After the pioneer experiments of Magendie (1783–1855), who fed to dogs only sugar, butter, gelatin and distilled water and found that they died in thirty to thirty-six days, the development of our present conception of the vitamins grew gradually out of repeated attempts, such as those of Lunin in 1881, to keep animals healthy on diets consisting entirely of purified foods. The idea of 'accessory factors of the diet' is strikingly illustrated in a series of experiments carried out by Hopkins between 1906 and 1912 (Fig. 10, 1). He showed that young rats fed on diets consisting of purified casein, starch, cane-sugar, lard and inorganic salts invariably ceased to grow in a very short time and finally became ill and died. If the purified diet was supplemented with a small daily allowance of milk, not more than 4 per cent. of the total food eaten, the health of the animals was maintained and they began to grow again. About the same time Stepp in Germany came to a similar conclusion which later experiments with highly purified or synthetic diets have amply confirmed.

Long before studies of the nutritional requirements of the growing animal were attempted in the laboratory, investigations into the cure or prevention of certain diseases, now recognized as being the result of a faulty diet, had been leading to a similar notion. Evidence that scurvy could be cured by the administration of the "fresh fruits of the earth and greens" had been available for centuries, and as long ago as 1753 Lind had proved that scurvy could be cured by eating oranges or lemons. Takaki (1887) had shown that the health of men in the Japanese Navy, who suffered greatly from beri-beri, could be markedly improved by increasing the proportions of meat and vegetables in the diet, and in 1897 Eijkman reported that beri-beri in fowls was the result of a diet restricted to rice from which the outer parts of the grain had been removed (polished rice), and that it could be cured by the administration of either rice polishings or of aqueous or alcoholic extracts of them.

The striking analogy between the 'accessory food factors' necessary for the health and growth of experimental animals and the unknown substances which were essential for the cure or prevention of beri-beri and scurvy led Funk (1911) to the hypothesis that 'vitamines,' compounds of unknown composition which he believed contained amino-nitrogen, are normally present in food in minute quantities and that their presence is essential to the complete well-being of the organism; conversely that their absence from the diet leads to illness and death. We now know that few of these accessory factors actually contain amino-nitrogen and therefore Funk's original word 'vitamine' has been shortened to 'vitamin.'

The mode of action of some of the vitamins is still unknown but in certain respects they resemble the hormones. For example, although quite small amounts are essential for health, continued failure of the supply is followed by disordered metabolism and finally by deficiency disease. The idea that lack of growth, loss of weight, severe disease and even death can be due to the lack of a single dietary material does not seem strange to us to-day but such a conception was accepted with great difficulty even fifty years ago. This apparently negative doctrine lacked the convincing power of bacteriology, at that time advancing rapidly, which attributed disease to positive agencies such as bacteria and protozoan parasites.

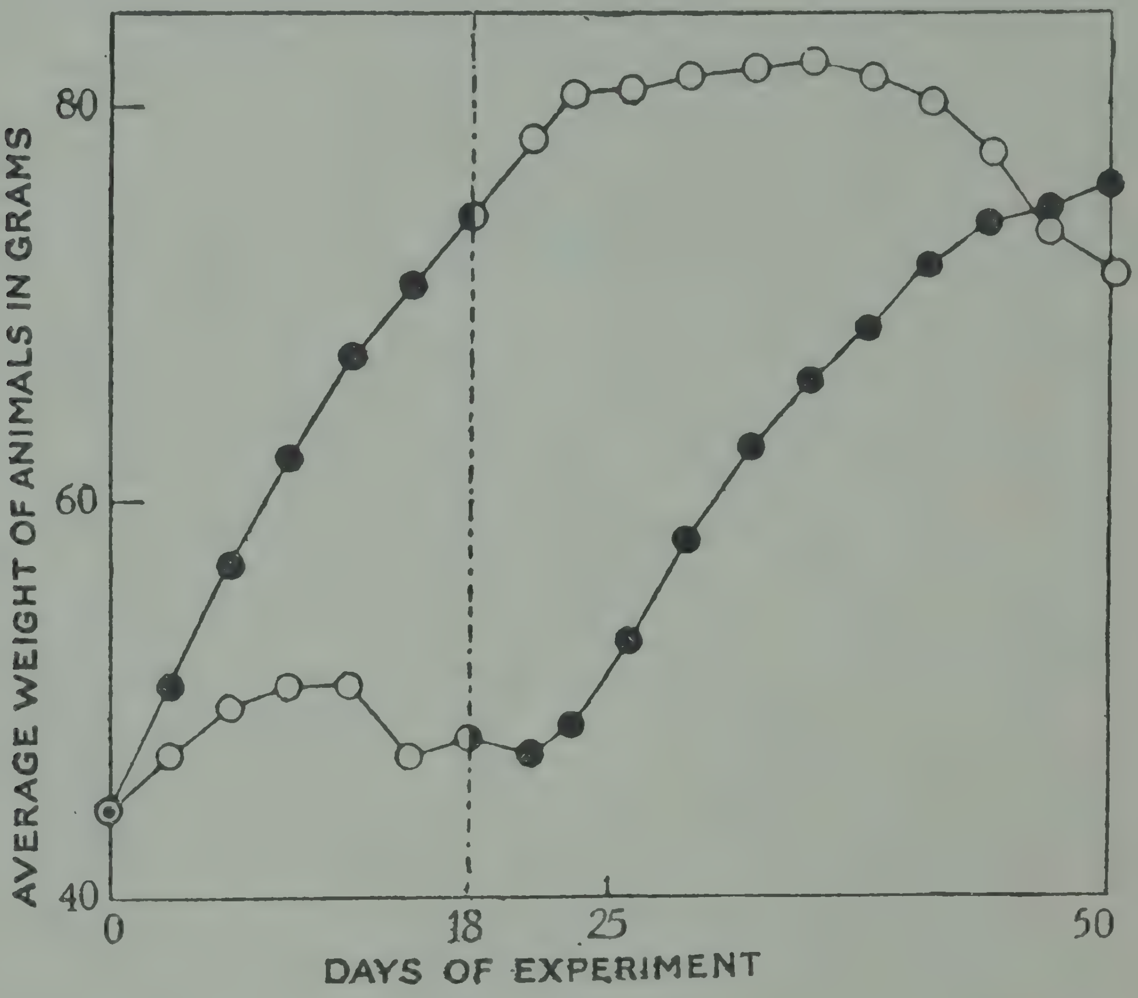


Fig. 10, 1

An experiment of Hopkins illustrating the conception of accessory food factors. "Lower curve (up to 18th day) eight male rats upon pure dietary; upper curve eight similar rats taking 3 ml. of milk each day. On the 18th day, marked by vertical dotted line, the milk was transferred from one set to the other." (F. G. Hopkins (1912), *J. Physiol.* **44**, 425.)

The next important step in our knowledge of the vitamins came in 1913, when McCollum and Davis showed that rats living on partly purified diets could be made to grow by the administration of ether extracts of egg-yolk or butter, but not by giving extracts of vegetable oils. Later experiments in 1915 showed that purified diets containing the fat-soluble growth factor must be supplemented by a water-soluble substance to promote growth. It was therefore clear that at least two factors, one fat-soluble and the other water-soluble, were required. Subsequent studies showed that each 'factor' was actually a group of vitamins a few of which even now have not been completely identified (Fig. 10, 2). When it became clear that both a fat-soluble factor and a water-soluble factor were necessary to supplement purified diets, the fat-soluble vitamin was labelled 'A' and the water-soluble vitamin was labelled 'B.' Later it was found that the material required for the prevention and cure of scurvy also came into the category of water-soluble factors and it was therefore given the letter 'C.' The discovery that the vitamin required for the prevention and cure of rickets was fat-soluble led to the name vitamin 'D.' The modern tendency is to abandon this alphabetical nomenclature in favour of the chemical name of the pure substances in cases where the structure of the material is known. Thus we tend nowadays to refer to ascorbic acid rather than to vitamin C.

The following pages contain brief descriptions of the individual vitamins and their modes of action so far as these are known at present. Although symptoms of vitamin deficiencies will be described in the sections dealing with individual vitamins, it must not be assumed that deficiency disorders such as beri-beri and pellagra are due to the lack of a single vitamin. For experimental purposes it is frequently possible to devise a diet lacking a single factor, but outside the laboratory a dietary deficiency is unlikely to be restricted to one substance, so that diets which produce vitamin-deficiency disorders lack not only a number of vitamins but may be deficient also in protein, essential amino-acids and minerals. Thus the results of feeding human volunteers an experimental diet which lacks only thiamine are very different from the disease beri-beri which is the consequence of a deficiency of thiamine modified by the lack of members of the vitamin B complex and perhaps other factors. Lack of a vitamin is not the only factor responsible for the production of symptoms. In starvation the features of vitamin deficiency disorder to be described presently are rarely found; it seems then that a certain level of nutrition or metabolic turn-over must be present before signs of deficiency appear. Furthermore, there may be variations in the individual response to deficiency since a diet lacking in vitamins of the B complex may produce beri-beri in one person and pellagra in another.

One of the commonest effects of an experimental deficiency of almost any vitamin in an immature animal is a retardation of growth, which becomes normal again when the missing vitamin is given. The efficacy of one vitamin in restoring growth in a deficient animal may be influenced by the simultaneous administration of other vitamins. An interrelation between the effects of one vitamin and another has also been demonstrated during the treatment of disorders due to deficiencies of the B-complex vitamins in man; it has been found, for example, that the administration of large doses of nicotinic acid amide may actually precipitate the signs and symptoms of riboflavin deficiency.

Although the ways in which normal metabolism becomes disturbed in deficiency are as yet largely unknown, such knowledge as we already possess suggests that the vitamins form essential links in metabolic processes. Riboflavin, for example, is incorporated into the prosthetic

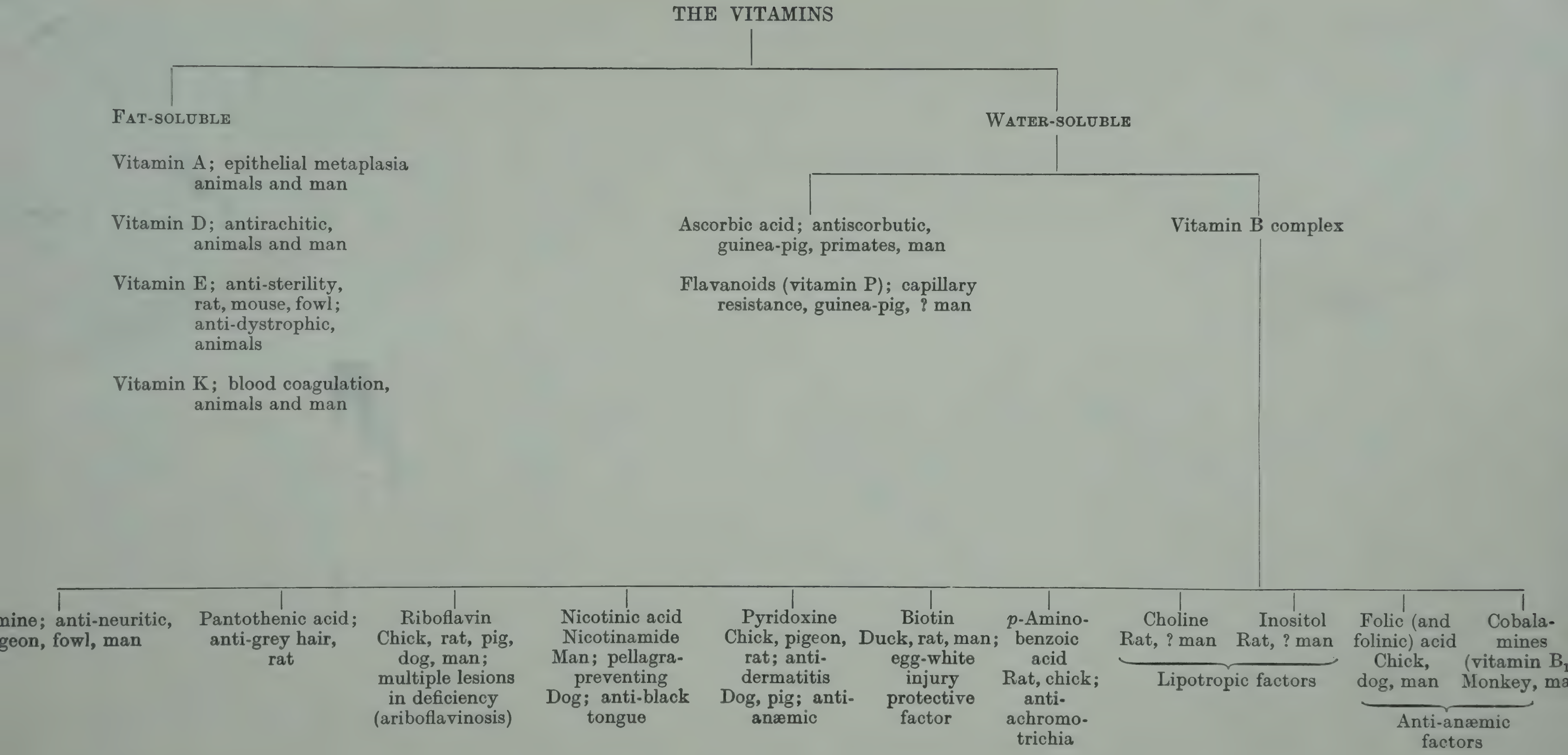

THE VITAMINS
FAT-SOLUBLE
Vitamin A; epithelial metaplasia animals and man
Vitamin D; antirachitic, animals and man
Vitamin E; anti-sterility, rat, mouse, fowl; anti-dystrophic, animals
Vitamin K; blood coagulation, animals and man
WATER-SOLUBLE
Ascorbic acid; antiscorbutic, guinea-pig, primates, man
Flavanoids (vitamin P); capillary resistance, guinea-pig, ? man
Vitamin B complex
nine; anti-neuritic,
geon, fowl, man
Pantothenic acid; anti-grey hair, rat
Riboflavin Chick, rat, pig, dog, man; multiple lesions in deficiency (ariboflavinosis)
Nicotinic acid Nicotinamide Man; pellagra-preventing Dog; anti-black tongue
Pyridoxine Chick, pigeon, rat; anti-dermatitis Dog, pig; anti-anæmic
Biotin Duck, rat, man; egg-white injury protective factor
p-Aminobenzoic acid Rat, chick; anti-achromotrichia
Choline Rat, ? man
Inositol Rat, ? man
Lipotropic factors
Folic (and folinic) acid Chick, dog, man
Cobalamines (vitamin B
Monkey, ma
Anti-anæmic factors

FIG. 10, 2

This table summarizes the protective actions of the vitamins and indicates the species in which effects of deficiency have been studied.

group of an enzyme which is essential for the transport of hydrogen. A dietary deficiency of riboflavin is ultimately revealed, therefore, by a disturbance of oxidative processes. In other words, the abnormality induced by riboflavin deficiency is in the first instance a disorder of chemical reactions in the tissues. It is only when such a functional disturbance (*biochemical lesion*) has been present for some time that structural changes can be recognized. Again, although pellagra is a disease in which characteristic structural alterations occur, it should be thought of primarily as a metabolic disorder in which many biochemical processes have either broken down completely or else have been continuing at reduced rates determined by inadequate supply of a number of essential dietary factors. For these reasons naturally occurring vitamin-deficiency disorders can only be treated effectively by the administration of all the missing factors, that is, by a fully balanced diet. It is not until gross structural changes have occurred that the administration of specific vitamins or, more logically, of a fully adequate diet is unable to restore the condition to normal.

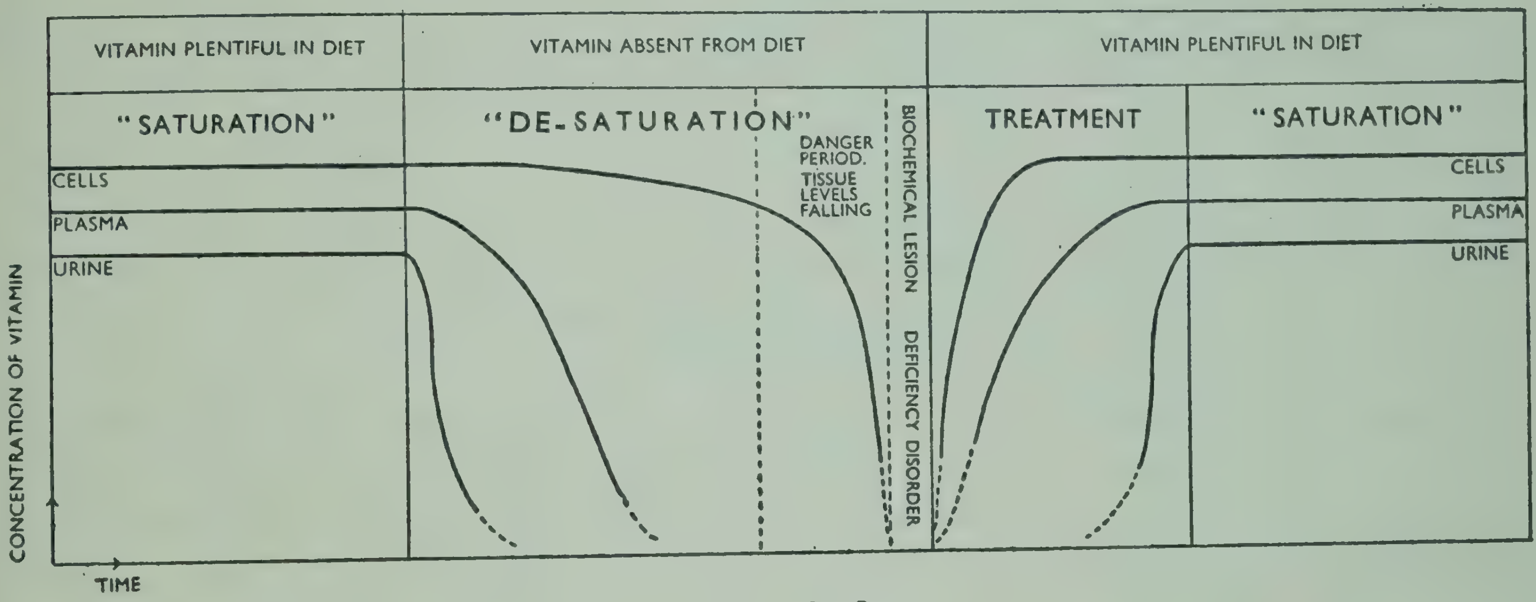


FIG. 10, 3

Accurate knowledge of the ways in which the several vitamins function in the animal can only be gained through reliable methods for measuring the concentration of vitamins in body fluids and tissues. Quantitative methods for estimating some of the vitamins have been devised and their application has led to the concept of *saturation* (Fig. 10, 3). The body is 'saturated' when the tissues are replete with the vitamin; the blood concentration is then steady at a high level and in equilibrium with that in the tissues and the daily excretion in the urine is comparatively constant. If the daily intake of the vitamin is now considerably raised, or if a single large dose is administered, there is little increase, except perhaps very temporarily, in either the tissue or blood levels, the extra vitamin appearing in the urine. Should the subject be given a diet from which the vitamin is lacking the amount of vitamin excreted in the urine falls rapidly to very low levels and the concentration in the blood also diminishes. It is not until somewhat later, the period being different for different vitamins, that the tissue concentration falls sufficiently low to disturb metabolic processes and ultimately to bring the patient to the brink of vitamin deficiency disease. These changes can be reversed by treatment with large doses of the appropriate vitamin and the subject can be saturated again. The state of saturation may then be maintained by smaller doses. At any time during a period of 'de-saturation' (Fig. 10, 3) the body's reserve store can be roughly assessed by measuring the effect of a single large dose of a vitamin on its excretion in the urine. The amount to which the urinary excretion is raised, or the percentage of the test dose which is excreted, may then serve as an index of de-saturation. Alternatively the amount of the vitamin which has to be given to restore saturation may be determined. Such tests, which are used clinically, are called 'saturation tests.' They cannot be applied to every vitamin because, for example, the fat-soluble vitamins are not excreted in the urine. Although saturation is not essential to perfect health such tests have provided information about vitamin requirements and have proved of value in the study of disease.

The conception of saturation is useful because it leads directly to a consideration of daily requirements. The daily requirement is a convenient statistical conception in terms of which can be expressed the need of the body for a particular vitamin. It is clear, however, from what has already been said, that an animal does not require a dose of a given vitamin every day in order to function efficiently. Indeed, under natural primitive conditions both man and animals must be able to survive quite long periods of dietary inadequacy occasioned, perhaps, by adverse climatic conditions, and it has been found for example with ascorbic acid that appropriate doses can be given at quite long intervals to guinea-pigs without any apparent disadvantage. Experiments on human volunteers have shown that the time elapsing between complete deprivation of a single vitamin and the appearance of symptoms varies considerably for different vitamins. This period is two to three months for thiamine and six to nine months for ascorbic acid, but it obviously depends on the degree of saturation of the subject at the beginning of the experiment. One may consider, therefore, the body's requirements for a given vitamin either in terms of the amount which must be taken daily in order to maintain saturation (*the optimal requirement*), or as the daily quantity which is just sufficient to prevent signs of disordered metabolism (*the minimum requirement*). The daily intake of a vitamin necessary for perfect health, while presumably greater than the minimum requirement is not necessarily as large as the optimal requirement. Assessments of vitamin requirements of man are in general based upon the following considerations:

(*a*) analogy with the results of animal experiments;
(*b*) studies of the vitamin content of human dietaries;
(*c*) the effect of graded doses of the vitamins in curing or preventing deficiency symptoms;
(*d*) blood levels, excretion studies and the results of saturation tests;
(*e*) special tests based upon the specific functions of certain vitamins (for example studies of dark adaptation for vitamin A).

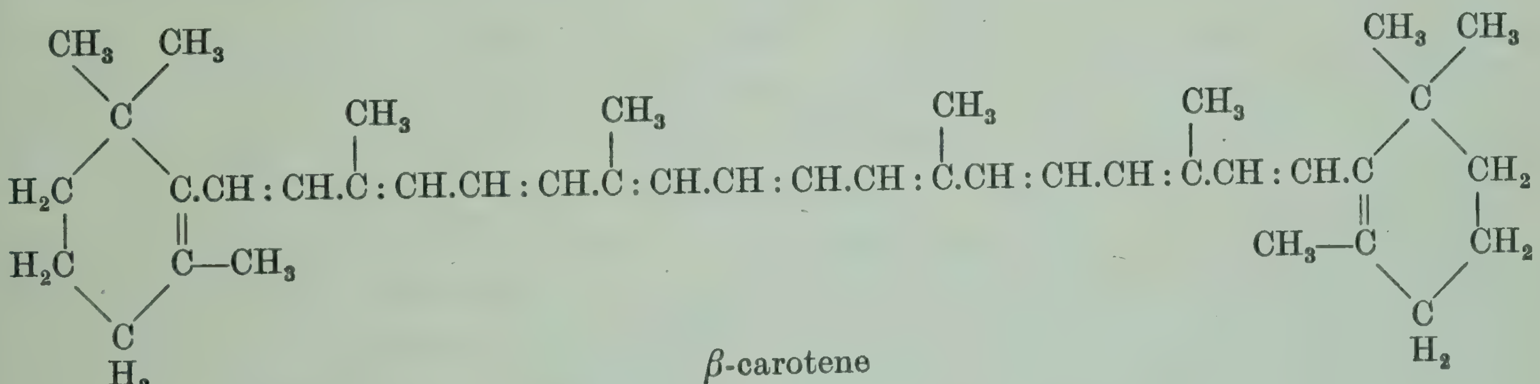
Since it is difficult at present to give optimal and minimal figures for any of the vitamins with confidence, it is more usual to speak of the 'average' daily requirement which lies somewhere between these two extremes.

The daily requirements for many of the vitamins are influenced by a number of factors. They are, for example, relatively greater in the growing child, during pregnancy and lactation, or when the metabolic rate is increased by exercise or fever. Disease of the intestine or of the liver may reduce the absorption of vitamins, particularly those of the fat soluble group, and liver disease may interfere with storage and metabolism. In the intestinal tract of man the bacterial flora normally synthesizes considerable quantities of vitamins, especially of members of the B complex, and in this way an important proportion of the daily requirement may be provided. Alteration in the flora, produced for example by the administration of certain drugs and antibiotics, may remove this source of supply.

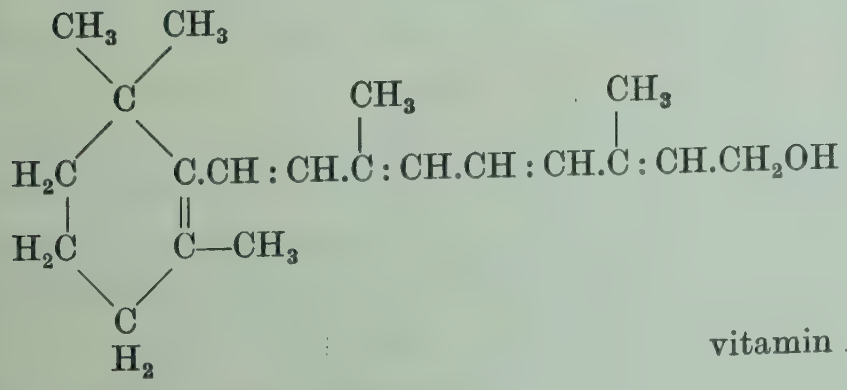
A guide to daily allowances of many of the vitamins is given in Table 12, 17, which was compiled by the Committee on Nutrition of the British Medical Association (1950). Lists of some of the more important foods from which the daily requirements may be met are given in this chapter. It must be emphasized that if the general principles for the construction of a diet given in Chapter 12 are followed, and if a wide variety of foodstuffs is eaten, a healthy person should have no difficulty in obtaining an adequate supply of all the vitamins.

VITAMIN A

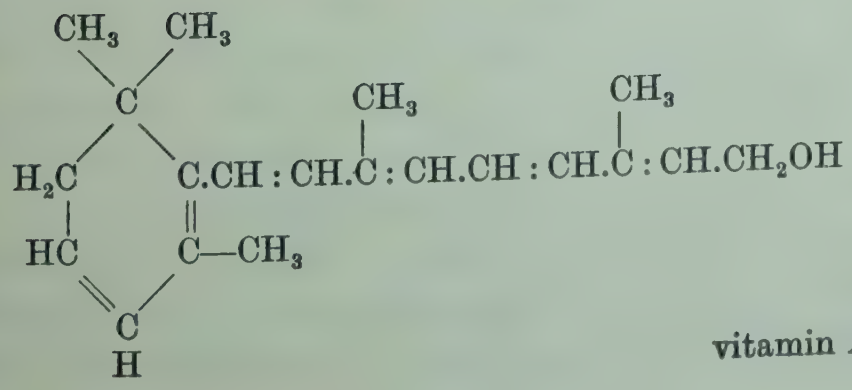
Although some of the symptoms produced in man by a deficiency of vitamin A have been known for centuries, the first steps in the discovery



β-carotene



vitamin A



vitamin A_2

of this vitamin were made in 1909 by Hopkins in Britain and by Stepp in Germany. Between 1913 and 1917 Osborne and Mendel, and McCollum and Davis in America differentiated 'fat-soluble A' which cured nutritional eye disease in animals and was necessary for the growth of rats. Vitamin A is now known not to be a single substance; there are at least three closely related chemical compounds which possess vitamin A activity. Plants do not contain pre-formed vitamin A but precursors of the vitamin are synthesized by plants and found particularly in the yellow and green parts. Animals cannot manufacture vitamin A except from a limited number of precursors, which they must therefore obtain from plants, and the vitamin itself, though not the precursors, is found in the organs of many species of animal in the form of esters.

Vitamin A precursors all belong to the class of carotenoids, hydrocarbons each with some forty carbon atoms, soluble in fat and fat solvents but not in aqueous solutions. These carotenoids are therefore called *provitamins*. The most important member of this group, β-carotene, was isolated from carrots as long ago as 1831, but its biological importance was not realized at that time. Comparison of the formulæ for β-carotene and vitamin A makes it seem possible that the β-carotene molecule might be symmetrically split into two molecules of vitamin A. In the body, however, the yield is less than this, progressively falling, at any rate in rats, as the intake of carotene is increased. Carotene is poorly utilized in children for whom it is very inferior to the vitamin as a dietary supplement.

Vitamin A_2, which is chemically closely related to Vitamin A, is found in the liver oils of fishes, especially those living in fresh water. Its functions are similar to those of vitamin A which it replaces in these species. Neovitamin A, a stereo-isomer of vitamin A, is also found in many fish-liver oils. It is probably converted in the body into vitamin A and has the same biological activity in rats.

Vitamin A was first synthesized in 1942 but in 1937 it had been isolated from the unsaponifiable fraction of fish-liver oil. In an inert atmosphere it is quite stable to heat but in the presence of oxygen it is easily oxidized to biologically inactive compounds. Vitamin A is slowly destroyed in fats and oils even at room temperatures, especially if the fat becomes rancid, but certain oils, for example wheat-germ oil, contain 'anti-oxidants' which inhibit the oxidation of vitamin A. Preparations of the vitamin should be kept in dark containers because vitamin A is rapidly destroyed by ultraviolet light. The international conference held in London in 1949 under the auspices of the World Health Organization recommended the adoption of vitamin A acetate as the reference standard and defined the unit of vitamin A activity as that shown by 0·344 μg. of pure vitamin A acetate. Vitamin A has a characteristic absorption spectrum which is sometimes used for its estimation, and its fluorescence in ultraviolet light has been used to detect its presence in animal tissues. Both vitamin A and the carotenoids give a blue colour with antimony trichloride in chloroform solution (Carr-Price reaction) and this reaction forms the basis of a method for their quantitative estimation.

The absorption of provitamins A, and of vitamin A itself, from the alimentary tract is favoured by the simultaneous absorption of fat or oil, especially if these are unsaturated, and conversely absorption of the vitamin is defective when the absorption of fat is impaired, as it is in sprue and cœliac disease. The ingestion of unassimilable oil such as liquid paraffin may likewise seriously interfere with absorption of the vitamin. Since vitamin A is absorbed only as the free alcohol its absorption depends upon lipolytic enzymes. Bile salts are also necessary for the absorption both of vitamin A and carotene. Small amounts of vitamin A are present in the body fat, but when the intake is high large amounts of the vitamin are stored in the liver; the liver of the polar bear contains so much vitamin A and related compounds that it is unfit for human food because of its toxicity. The conversion of the provitamins into vitamin A occurs mainly in the intestinal mucosa.

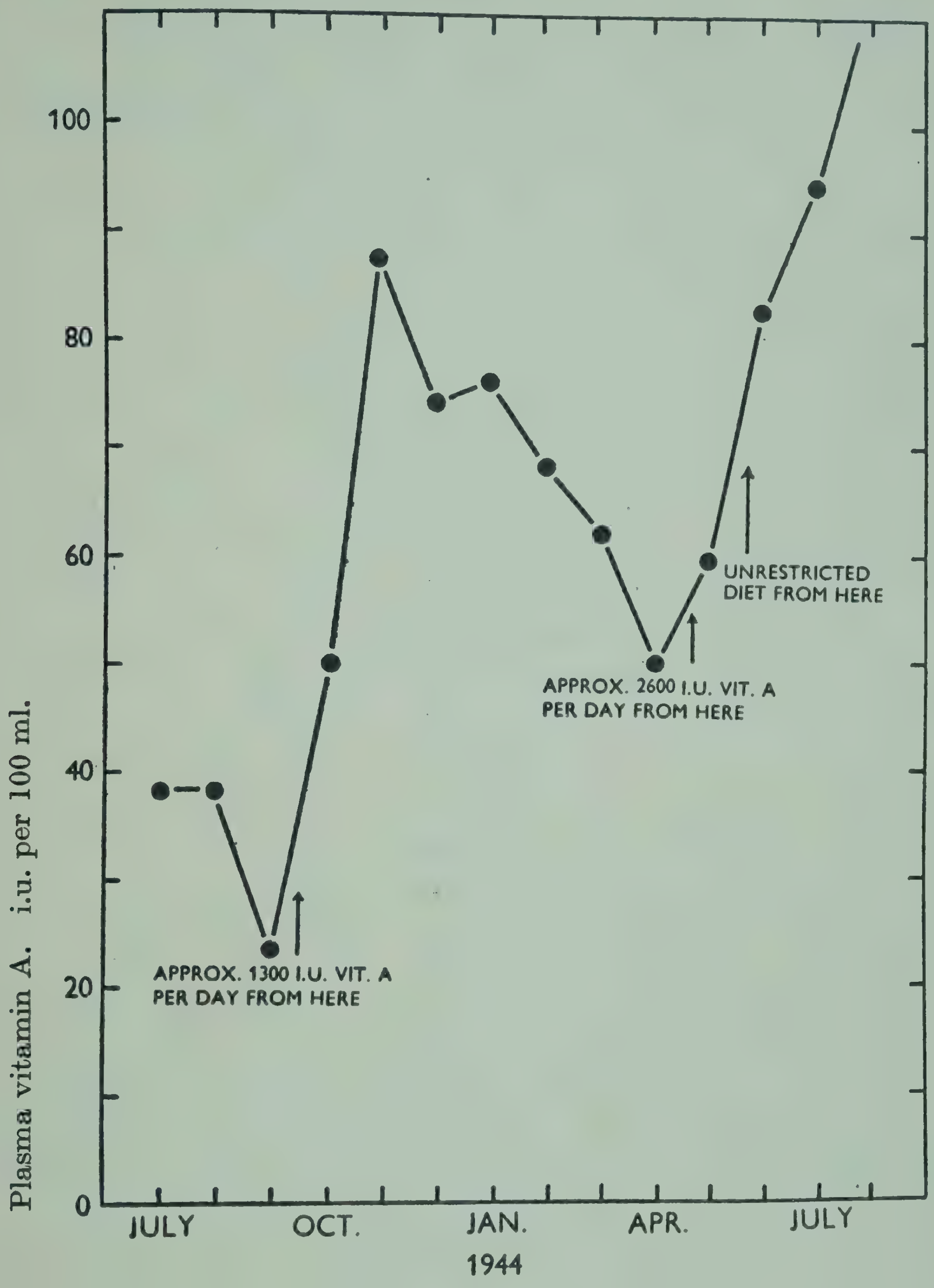


Fig. 10, 4

Showing the low vitamin A concentration in the plasma in a subject depleted of vitamin A and its gradual increase during administration of the vitamin (Vitamin A requirements of human adults (1949). *Spec. Rep. Ser. med. Res. Coun., Lond.* No. 264. London: H.M.S.O.).

Both vitamin A and carotene are present in blood plasma, the former probably linked to the plasma protein; the concentration of vitamin A in health is from 100 to 130 i.u. per 100 ml. and of total carotenoids 120 to 180 i.u. per 100 ml. These values fall during depletion but normal figures for vitamin A are restored by administration of the vitamin (Fig. 10, 4).

Vitamin A deficiency may therefore result from inadequate intake or from defective absorption of the vitamin or provitamins, or from interference with the conversion of provitamins into vitamin A. Since the biochemical lesion of vitamin A deficiency is unknown we are limited to a description of its effects, which are to be found in the epithelial tissues derived both from endoderm and ectoderm, namely, the mucous membranes and the skin together with their associated glandular structures, the retina and the central nervous system. In mucous membranes the earliest and most characteristic change is a gradual replacement of the normal epithelium by a stratified keratinizing epithelium. This alteration of cell type, or *metaplasia*, leads to diminished secretion of mucus and to impairment of ciliary activity. In glandular alveoli and in the folds of the conjunctiva and mucous membranes the accumulation of keratinized debris and dried secretion encourages infection, which provokes local inflammatory reactions. These changes are particularly striking in the eye, producing first *xerophthalmia* (literally 'dry eye') and later keratitis and conjunctivitis (inflammation of the cornea and conjunctiva respectively). If the patient is not given vitamin A, softening, disorganization, and finally complete and permanent destruction of the cornea results (*keratomalacia*) (Fig. 10, 7). Vitamin A deficiency is accompanied by severe damage to both rods and cones (Fig. 10, 7). Metaplasia in the mucous membrane of the respiratory tract is followed eventually by infection which is often fatal. In man the skin becomes thick and dry, the sebaceous and sweat glands atrophy and the ducts of the hair follicles become blocked by plugs of keratin which produce an apparent roughness of the skin (Fig. 10, 8). This rough skin (*follicular hyperkeratosis*), however, is not specifically characteristic of vitamin A deficiency. In some animals receiving an inadequate supply of vitamin A osteoclastic activity is much reduced and the normal process of removal of bone from the internal surfaces of the skull and spinal canal is restricted, so that these cavities become insufficient to accommodate the growing brain and spinal cord. Compression of the nervous tissue produces disturbance of function, particularly in the cranial nerves. In young animals deprived of vitamin A degenerative changes in the tissues of the central nervous system have also been described.

It has been pointed out by Wald that in both animals and plants light-sensitive structures are always found to contain carotenoids. 'In this respect the appearance of vitamin A in the mammalian retina continues an association which extends so far as is known throughout the eyes of vertebrates and the photosensitive organs of multicellular plants, algæ and fungi.' An early sign of vitamin A deficiency in man is *night-blindness*. This is an inability to see properly in dim light and is especially obvious if the eyes have recently been subjected to bright light. It has, for example, been described as affecting fishermen after a day at sea exposed to bright sunlight which is strongly reflected from the surface of the water. As the result of the work of Morton, of Wald and of others it has been established that the retinal pigment rhodopsin (visual purple), a conjugated protein containing a carotenoid pigment as a prosthetic group, is essential for normal vision in dim light. It is bleached by bright light to a yellow pigment retinene (vitamin A aldehyde) and regenerated in the dark (Fig. 10, 5). Since the retinene produced by bleaching is not all available for resynthesis, the ability to see in dim light depends on the regeneration of rhodopsin from some other source. This source is the vitamin A in the plasma. The value of raw liver in relieving night-blindness has been known for centuries. By measuring the rate of dark adaptation (Chapter 39) the vitamin A status of an individual may be roughly assessed.

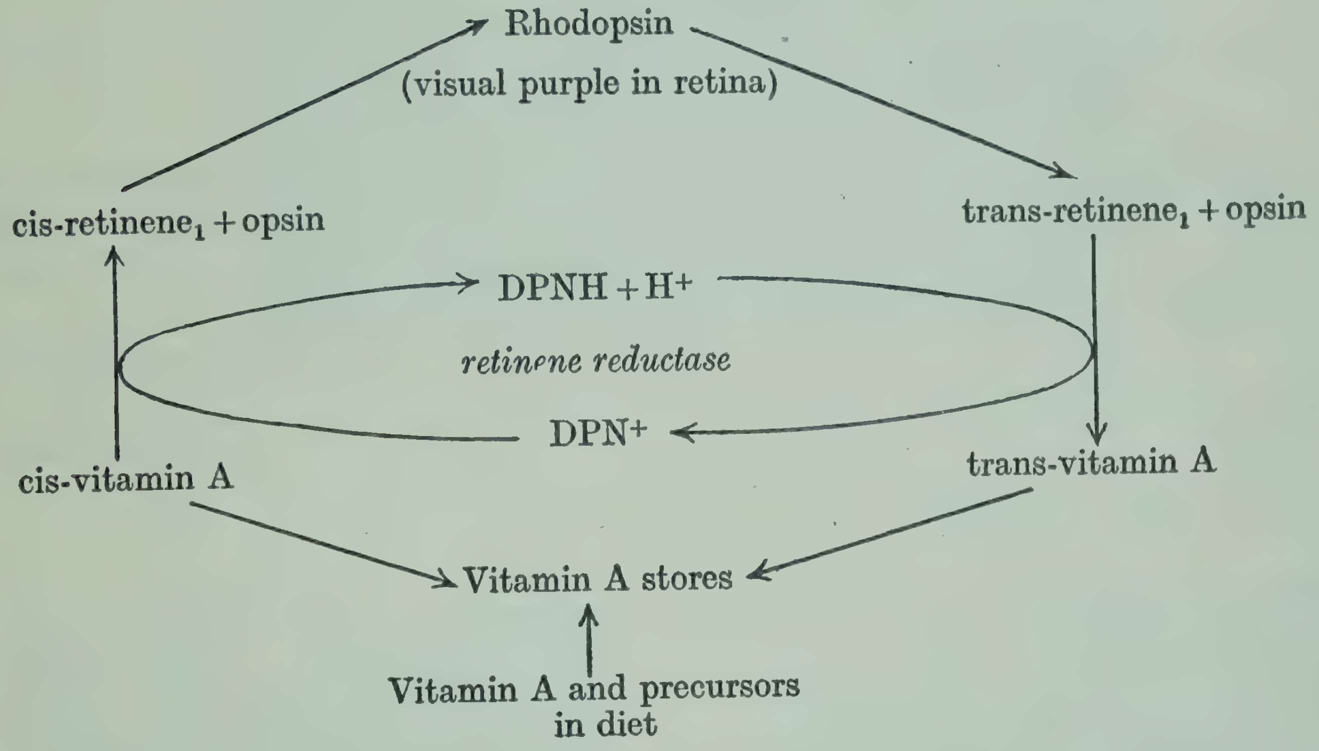


FIG. 10, 5

The visual cycle (after Wald).

The provitamins A, for example carotene, are converted in the intestinal epithelium to vitamin A (an alcohol) which, like its aldehyde retinene$_1$, can exist in stereoisomeric cis- and trans-forms, of which the former are the more active biologically. After absorption, vitamin A is stored as cis-vitamin A, mainly in the liver, but it is present in plasma in which it reaches the retina where it is fixed as rhodopsin. By the photo-chemical action of light on the retina rhodopsin is converted to trans-retinene$_1$ and the protein opsin; the retinene$_1$ is then reduced under the influence of the enzyme retinene reductase and DPN to trans-vitamin A. This passes via the circulation to the liver to be isomerized to cis-vitamin A for storage. It is liberated again from the liver into the circulation from which it is taken up by the retina to be converted by retinene reductase to cis-retinene$_1$. This in turn combines with opsin to form rhodopsin. In fresh water fishes vitamin A, rhodopsin and retinene$_1$ are replaced respectively by vitamin A$_2$, porphyropsin, and retinene$_2$.

Preformed vitamin A is found in comparatively few articles of a normal diet. The commonest foods containing nutritionally important amounts of vitamin A are eggs, milk, butter and vitaminized margarine, but it should be remembered that the vitamin A content of farm produce depends upon the diet of the cow or fowl, and therefore upon the season of the year; in Britain vitamin A is added to margarine to the amount of 760 to 940 i.u. per ounce (28 g.). Liver and liver oil contain relatively large amounts of vitamin A; fat fish (herrings, sardines, and salmon for example) are also good sources. The provitamins are widely distributed in vegetables and fruits, where they are found in association with chlorophyll and other pigments, so that in general the greener or more highly coloured parts are the best sources of the vitamin (Table 10, 6). It should be emphasized again, however, that to obtain the approximate vitamin A potency the carotene content of a foodstuff should be divided by a factor of between two and three. The preparation and cooking of vegetables impair their vitamin value relatively little, though deep-fat frying results in considerable losses, not only in the fat but also in materials cooked in it. Canning, dehydration and deep freezing may give rise to considerable losses. Neither vitamin A nor carotene is adversely affected by storage in the dark.

TABLE 10, 6

*Vitamin A Potency of Foods.** i.u. *per* 100 *g.*

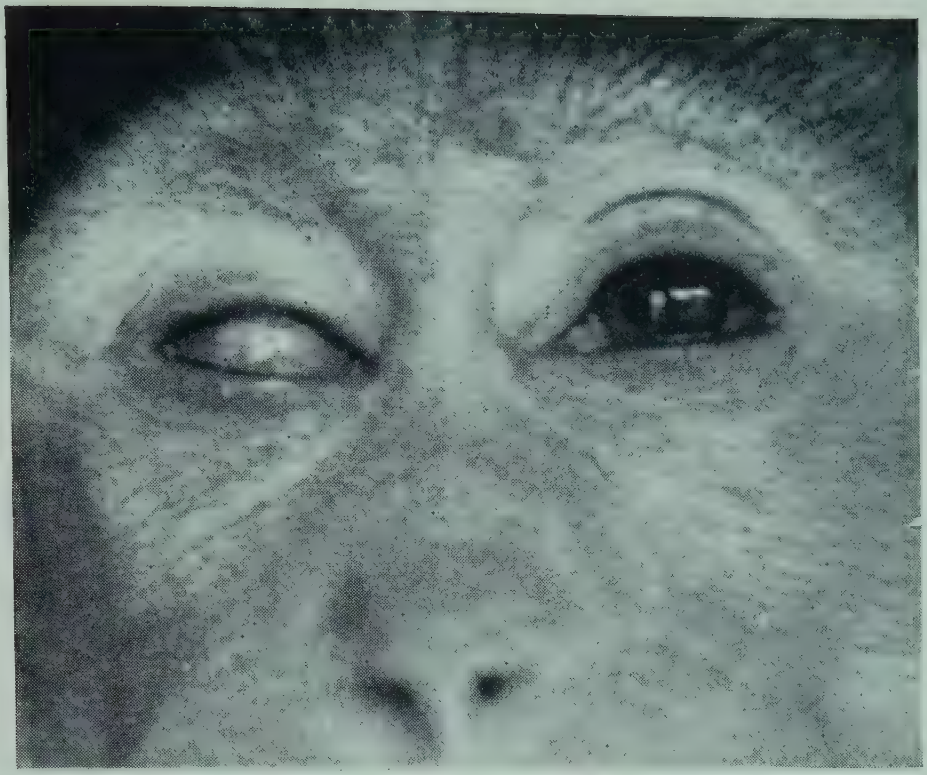
Halibut-liver oil	2,000,000 to 36,000,000
Cod-liver oil	40,000 to 400,000
Fresh herring	90 to 120
Canned herring	30
Ox liver	12,000 to 40,000
Beef and mutton (lean)	30 to 50
Egg (whole)	700 to 1000
Cow's milk † (raw)—summer	140
winter	70
Dried milk—whole	1225
skimmed	30
Human milk	200 to 500
Butter	2000 to 5000
Cheese (cheddar type)	1300
Margarine (vitaminized)	2300 to 3000
Carrots	10,000 to 16,000 ‡
Spinach	10,000 ‡
Kale	5600 ‡
Watercress	4250 ‡
Lettuce	3200 ‡
Tomatoes	2550 ‡
Cabbage	630 ‡

* The values in this and subsequent tables are mainly from "Nutritive Values of Wartime Foods" (1945). *M.R.C. War Memorandum No.* 14. London: H.M.S.O. The vitamin content of any individual sample may differ widely from the value found in the table.

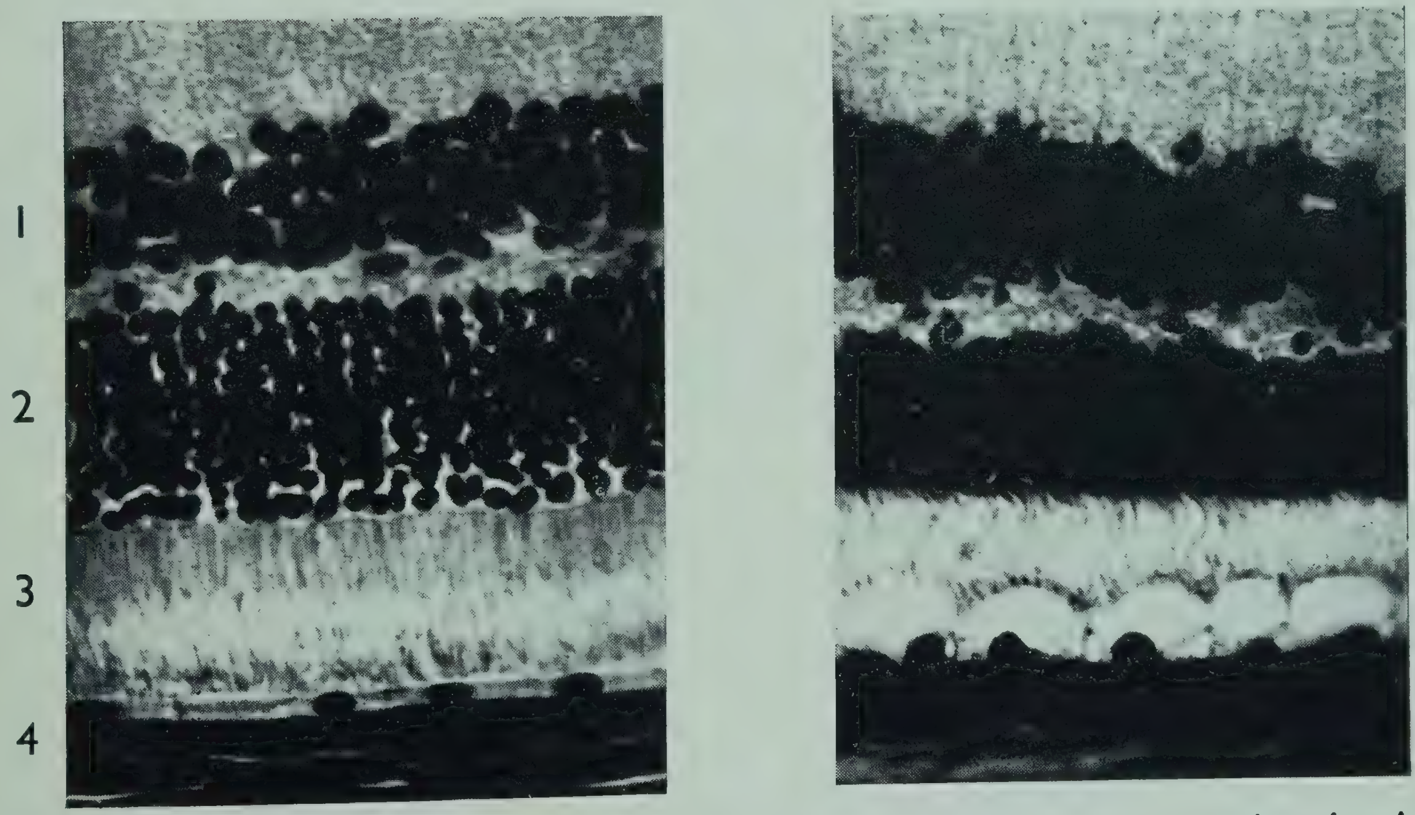
† Large variations in content are found according to breed of cow, yield of milk, pasture, season, etc.

‡ Potency is due to provitamin.

In human volunteers who were fed over long periods a diet containing only very small amounts of vitamin A and carotenoids, it was found that the minimum daily dose effective as protection against deficiency was about 1300 i.u. of vitamin A. A figure of 2500 i.u. per day of vitamin A for an adult has recently been recommended as an estimate of the requirement to cover individual variations and to leave a margin of safety (1950, Report of the Committee on Nutrition, British Medical Association). This dose is sufficient to sustain perfect health in a young, healthy male; whether it is large enough to allow accumulation of reserves is not known. The higher requirements during pregnancy and lactation are covered by recommendations of 3000 i.u. and 4000 i.u. per day respectively.



A. Keratomalacia of right eye in a monkey after 4½ months of a diet deficient in vitamin A.



B. Retina of control rat. Masson's trichrome stain. × 600.

C. Similar but from vitamin A deficient rat, showing atrophy of the rods and cones and swelling of the pigment cells.

Fig. 10, 7

V. Ramalingaswami, E. H. Leach & S. Sriramachari (1955), Ocular structure in vitamin A deficiency in the monkey. *Quart. J. exp. Physiol.* **40**, 337–347. 1. inner nuclear layer, 2. outer nuclear layer, 3. layer of rods and cones, 4. pigment epithelium. B. and C. are to be compared with Fig. 39, 12.

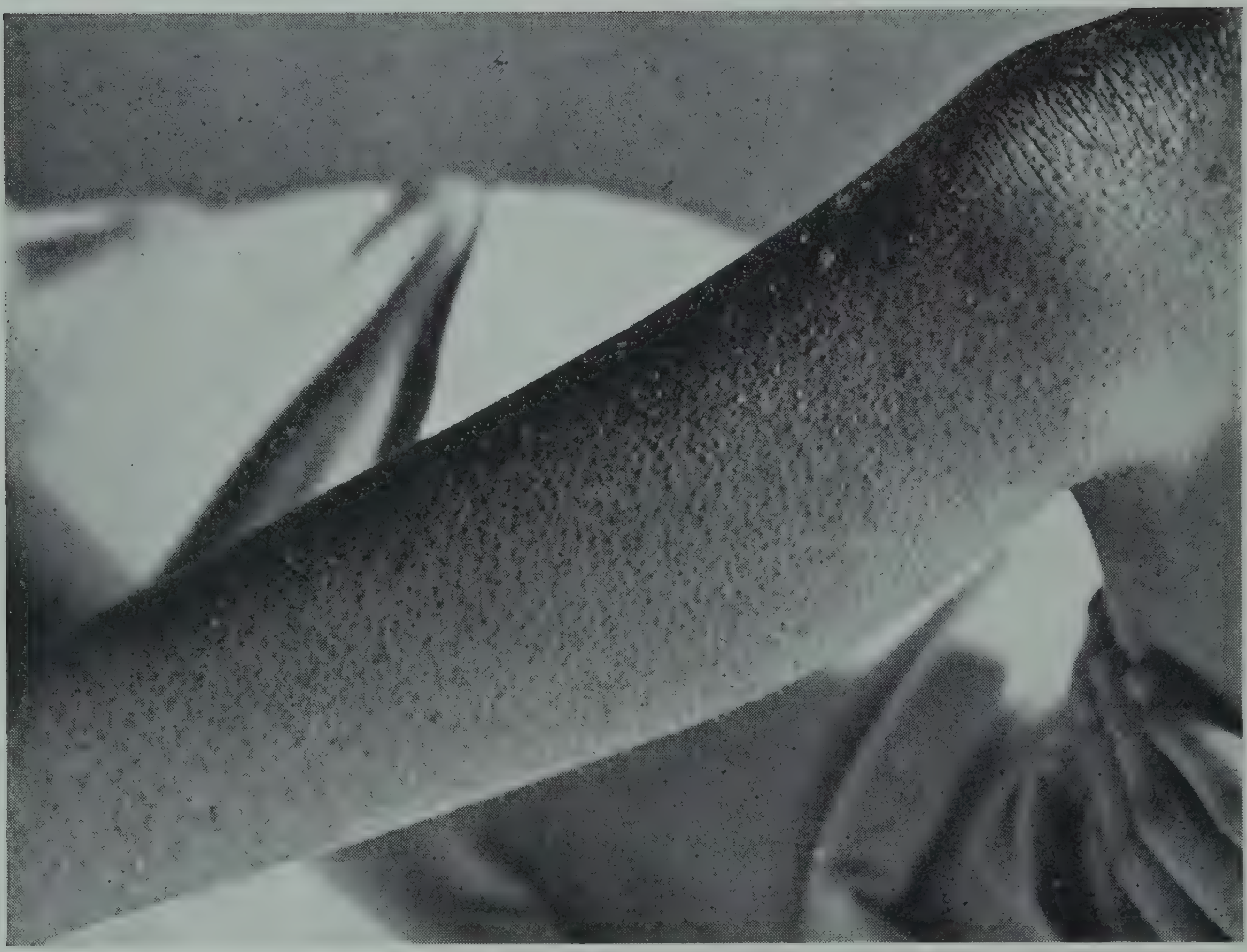


Fig. 10, 8

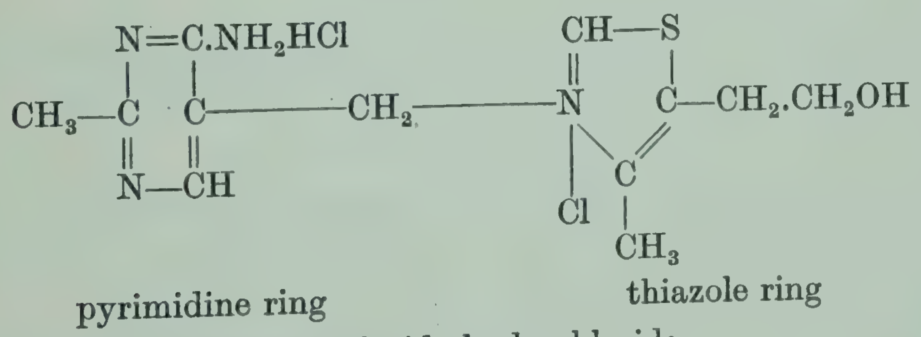
Follicular hyperkeratosis on the leg of a boy aged ten.

THE VITAMIN B COMPLEX

As mentioned earlier, McCollum and Davis in 1915 were led to postulate two groups of accessory food factors, 'fat-soluble A' and 'water-soluble B.' It was not until 1926, however, that vitamin B, as it was then called, was shown to consist of a heat-labile component having the properties formerly ascribed to vitamin B and a heat-stable component. These were called, respectively, vitamin B_1 and vitamin B_2, but it has since been shown that the vitamin B group is much more complex than this, and, although the name vitamin B_1 is still sometimes employed to designate the heat-labile substance thiamine, the term vitamin B_2 is not now used. About a dozen substances are now included in the vitamin B complex and the more important of these are described individually in the following pages.

THIAMINE (ANEURINE)

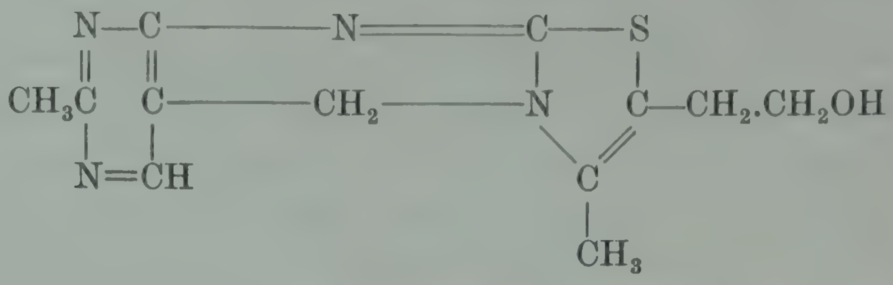
Our knowledge of the physiological properties of thiamine begins with the work of Eijkman with fowls, to which we have already referred. Between 1906 and 1909 his experiments were extended to man by Fletcher and by Fraser and Stanton, who proved that beri-beri could be produced by a diet of rice from which the outer coats (polishings) had been removed; they were able to cure the condition by giving the patients a similar diet containing the whole rice grain in place of the polished rice, thus showing that beri-beri is a deficiency disease. Funk in 1911 cured pigeons of a similar disorder by the administration of a highly active concentrate prepared from rice polishings, but it was not until 1926 that the pure vitamin was finally isolated. The synthesis of thiamine was accomplished in 1936.



thiamine chloride hydrochloride

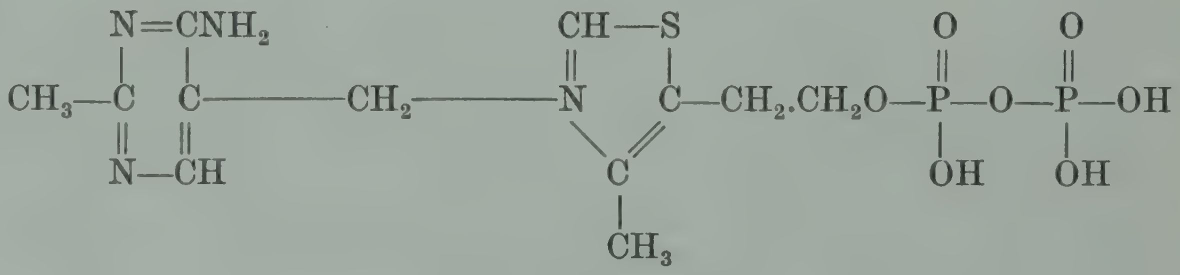
Thiamine, which consists of a pyrimidine and a thiazole ring system joined by a methylene bridge, is freely soluble in aqueous solutions. In acid solution it withstands prolonged heating at 120° C but, since thiamine is less thermostable than other members of the B complex, especially in alkaline solution, it forms, along with pantothenic acid, the thermolabile component of vitamin B. Under ordinary conditions thiamine is not oxidized by atmospheric oxygen but is readily converted by mild oxidizing agents to the pigment thiochrome which has no vitamin activity.

The concentration of thiamine in biological fluids and tissues can be measured by first converting the vitamin into thiochrome with potassium ferricyanide and then estimating the thiochrome by its bluish fluorescence in ultraviolet light. The growth of certain moulds, yeasts and bacteria is closely related to the concentration of thiamine in the nutrient medium and this forms the basis of biological assay.



thiochrome

The physiological actions of thiamine are due to its pyrophosphoric ester, co-carboxylase (thiaminepyrophosphate, TPP), which is formed by the phosphorylation of the vitamin under the influence of ATP and magnesium ions. Thiaminepyrophosphate acts as coenzyme in the decarboxylating enzyme systems which are involved in the body in the decarboxylation of such α-keto acids as pyruvic acid which is an important intermediate in carbohydrate metabolism (p. 314). In acting as a coenzyme thiaminepyrophosphate is frequently associated with lipoic acid (see below). In normal conditions pyruvic acid is metabolized by a series of reactions involving TPP (Chapter 20), but in thiamine deficiency the coenzyme is lacking and pyruvic acid accumulates in the body tissues and fluids.



thiaminepyrophosphate, TPP, or co-carboxylase

Animals maintained on a diet rich in carbohydrate but poor in fat develop signs of thiamine deficiency more rapidly than animals on diets with a high fat content and little carbohydrate. In man, too, the daily requirement for thiamine is related to the consumption of carbohydrate; a person eating a high carbohydrate diet requires more thiamine than one consuming mainly fat and protein.

It was demonstrated by Peters that lactic acid and pyruvic acid accumulated in the brains of pigeons suffering from thiamine deficiency and at the same time the oxygen consumption of the brain tissue was reduced. The latter was increased by the addition of thiamine. The importance of thiamine for the proper functioning of nervous tissue is also suggested by the fact that deficient animals regularly develop gross disturbances of function in the nervous system embraced by the name 'polyneuritis.' The functional derangement, of which the main features are muscular weakness and incoordination, is not necessarily associated with any structural change in the nerves and therefore deserves the name *neuropathy* rather than neuritis. The disorder may be the result of the deranged metabolism of carbohydrate or perhaps of a disturbance of the transmission of nerve impulses. The part played by acetylcholine in this process is described in Chapter 42. Thiamine has been claimed to augment the activity of acetylcholine by inhibiting the formation of cholinesterase.

The blood contains some 5 to 11 μg. of thiamine per 100 ml. almost entirely in the form of co-carboxylase. The total amount of the vitamin in the body is of the order of 25 mg.; the heart has the highest concentration (230 μg. per 100 g.), brain, liver and kidney contain about 100 μg. per 100 g. and skeletal muscle about 50 μg. per 100 g. As the body cannot store large quantities of thiamine the period of survival on a deficient diet is not significantly prolonged by prior administration of large amounts of the vitamin.

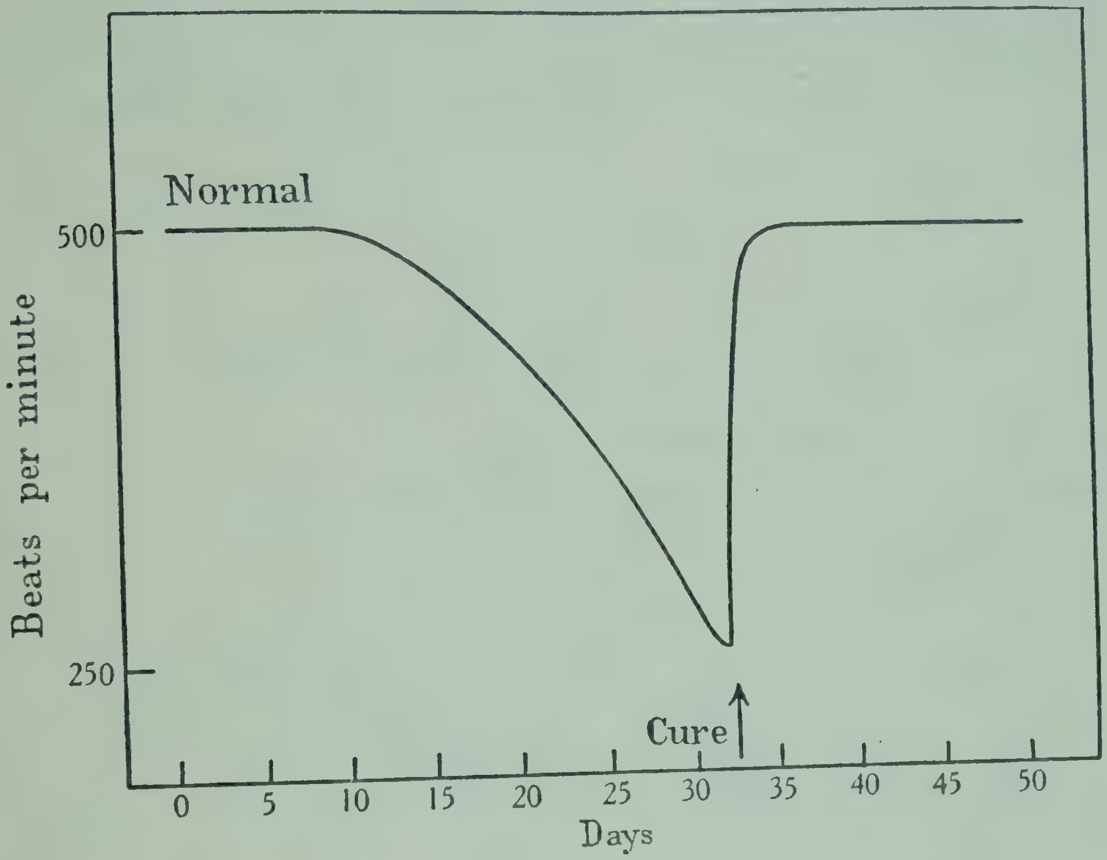


FIG. 10, 9

Showing the fall in heart rate (bradycardia) which occurs in a rat fed on a diet deficient in thiamine and the very rapid restoration of the normal rate when the vitamin is administered (L. J. Harris (1935), *Vitamins in Theory and Practice* (p. 66). Cambridge University Press).

Relatively large amounts of thiamine are found in the heart, and during conditions of depletion the vitamin is lost from cardiac muscle much more slowly than from other tissue. These facts suggest that the vitamin has an important role in cardiac metabolism, an idea which is supported by the finding that the cardiac rate is much below normal in pigeons and rats lacking this factor. The increase in heart rate produced by the administration of material rich in thiamine to a deficient rat is in fact the basis of a procedure, the *bradycardia test*, for assaying the thiamine content of foods (Fig. 10, 9). It is no doubt significant also that many of the earlier signs and symptoms of vitamin B deficiency in man are referable to the cardiovascular system, for example rapid and irregular heart action, cardiac enlargement and electrocardiographic abnormalities.

The disorder which results in man from a prolonged deficiency of vitamin B has for long been known as *beri-beri*. This condition, which is still common in many parts of Asia, is characterized by three groups of symptoms: (*a*) a disturbance of function in the peripheral nerves, both motor and sensory (polyneuropathy), which displays itself as a progressive muscular weakness followed by atrophy and a staggering gait (ataxia) (Fig. 10, 12), with inability to make purposeful movements of the hands and arms and with multiple disturbances of sensation culminating in anæsthesia; (*b*) cardiovascular symptoms terminating in severe cardiac failure; and (*c*) a rapidly progressing œdema (Fig. 10, 13). It is now becoming clear, however, that many of the signs and symptoms of beri-beri as it occurs in the East are the consequence of combined deficiencies of several vitamins of the B complex, and possibly also of other factors, rather than the result of thiamine deficiency alone. Although certain symptoms, such as lassitude, muscular weakness, loss of weight and disturbances of sensation, are common to beri-beri and the syndrome produced in volunteers deprived of thiamine only, œdema and cardiac failure have never been produced experimentally in man.

Table 10, 10

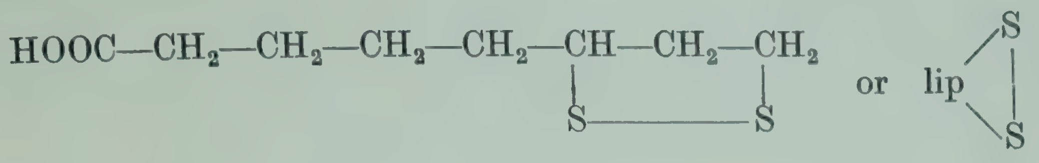
Thiamine in Foods. mg. per 100 *g.*

Wheat (germ)	2·5 to 5·6	Eggs	0·1 to 0·15
Flour, wholemeal	0·35 to 0·4	Beef and mutton	0·05 to 0·1
85 per cent. extraction	0·30 to 0·4	Pork	0·55 to 0·7
70 per cent. extraction (white)	0·05	Fish	0·02 to 0·06
Oatmeal	0·5	Fruits	0·02 to 0·05
Yeast	1·8 to 3·0	Fresh vegetables	0·02 to 0·1
Milk	0·04 to 0·06	Green peas	0·17
		Pulses	0·5 to 0·9

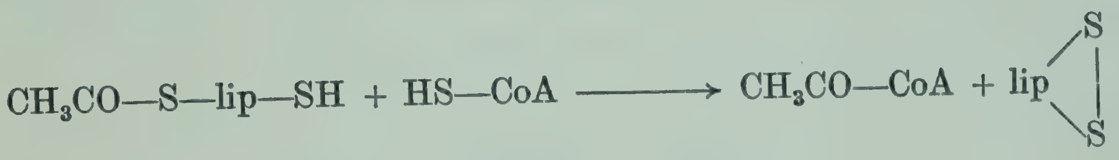
Thiamine is present in many plants, in particularly high concentrations in the seeds and the outer coats of grains (bran); the pulses and nuts are among the richest natural sources. White flour contains little thiamine because most of it is removed in the milling process. On the other hand wholemeal flour contains 350 to 400 μg. per 100 g. and is therefore a relatively good source. The 85 per cent. extraction flour (containing 85 per cent. of the whole grain) used in Britain between 1942 and 1944 contained 300 to 400 μg. per 100 g. The vitamin is widely distributed in animal tissues and the lean parts of meat, especially pork, may be important dietary sources (Table 10, 10). Yeast and yeast extracts contain very large amounts of thiamine and many bacteria are capable of synthesizing it. Although high temperatures and an alkaline medium favour destruction of the vitamin, losses during the preparation and cooking of food are not large. Freezing, canning and dehydration likewise result in small losses.

Lipoic Acid (Thioctic Acid).—Lipoic acid (pyruvate oxidation factor, protogen) is a growth factor for certain protozoa and micro-organisms.

It is a fat soluble material containing eight carbon and two sulphur atoms and hence is also known as thioctic acid.

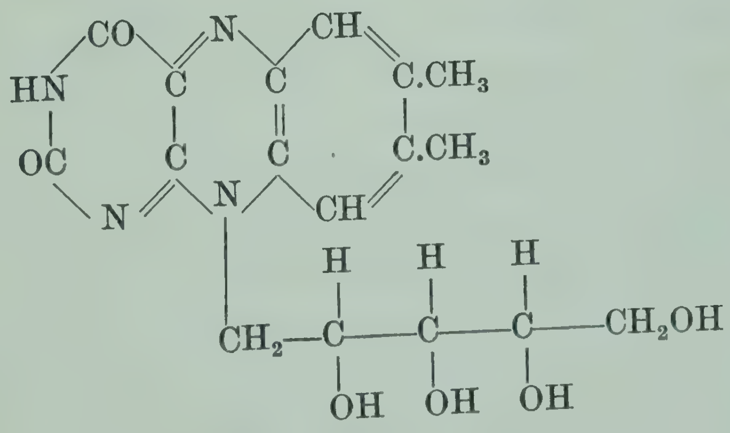


While decarboxylation reactions in general require TPP, some also require the presence of lipoic acid. Gunsalus has suggested, for example, that in the decarboxylation of pyruvic acid, $CH_3COCOOH$ (Chapter 20), the first step is the removal of carbon dioxide with the formation of the TPP derivative CH_3CO-TPP. This in turn reacts with lipoic acid to form an acetyl derivative in which the acetyl group is attached to one of the sulphur atoms thus: CH_3CO-S-lip-SH. This product now reacts with coenzyme A (p. 62) to form acetyl coenzyme A thus:—



RIBOFLAVIN

Although Smith and Hendrick showed in 1926 that water-soluble vitamin B consisted of at least two factors, it was not until 1932 that Warburg and Christian isolated from yeast a 'yellow enzyme' which they later showed could be separated into a protein component and a coloured substance neither of which was active without the other. The yellow compound proved to be riboflavin phosphate. Riboflavin itself was finally synthesized in 1935.



riboflavin (6 : 7-dimethyl-9-D-ribityl-isoalloxazine)

The yellow fluorescence exhibited by riboflavin in ultraviolet light forms the basis of a chemical method for its quantitative estimation. A microbiological method has been developed which depends on the fact that the growth of a strain of *Lactobacillus casei* and the resulting production of acid is proportional to the concentration of riboflavin in the culture medium.

A number of synthetic flavins possess in very mild degree the biological properties of riboflavin, but the latter is the only naturally occurring flavin which is found in the flavoprotein enzyme systems either as a mononucleotide, riboflavin phosphate, or as riboflavin-adenine-dinucleotide. The role of the flavoproteins in biological oxidations is dealt with in detail in Chapter 8. Here it is sufficient to note that the flavoproteins act as reversible carriers of hydrogen in oxidation processes and that some of the most important members of the group act as intermediates between the pyridine nucleotides and the cytochrome system.

Riboflavin is readily absorbed from the upper alimentary tract. It is present in blood mainly in combination with plasma globulin to the extent of some 0·5 μg. per 100 g. High concentrations of the vitamin are found in the liver, heart and kidneys, which retain relatively high amounts even when the body is seriously depleted. Riboflavin is eliminated in the urine in the form of a pigment, uroflavin, the excretion of which fluctuates with the intake of the vitamin. Riboflavin must be phosphorylated before it acquires biological activity. This probably occurs in the cells of the intestinal mucosa and in the liver.

In view of the key part played by the riboflavin molecule in several enzyme systems it is remarkable that the symptoms of deficiency in man are not more dramatic. Although they were described by Stannus in 1911, many years before riboflavin was discovered, the symptoms were not generally recognized until after Sebrell and Butler's description in 1938 of experimental deficiency of riboflavin in man. The syndrome, called *ariboflavinosis*, involves particularly the eyes and skin, especially at junctions between the skin and mucous membrane. The abnormality of the skin (dermatosis) is often best seen in the face, where the skin is roughened, scaly and greasy. The pink parts of the lips become bright red, swollen and cracked (cheilosis) and the tongue enlarged, tender and magenta in colour (Fig. 10, 14A). The lesions of mucocutaneous junctions occur characteristically at the corners of the mouth, which are sore and fissured (angular stomatitis) (Fig. 10, 14B), but similar lesions may be present around the anus and vagina. In both sites desquamation of the skin occurs and the tissues readily become infected. The ocular manifestations of ariboflavinosis include congestion of the conjunctival blood vessels, especially those encircling the cornea (circumcorneal injection) and of redness and swelling of the eyelids (blepharitis). Minor grades of riboflavin deficiency can be detected by microscopic examination of the patient's eye under a slit-lamp: the blood vessels at the margins of the cornea are proliferated and the transparent corneal tissue is gradually infiltrated by loops of capillaries growing in from the vessels surrounding the cornea (limbic vessels). When the deficiency is more advanced the subepithelial layers of the cornea are infiltrated by leashes of new blood vessels spreading in from the edge and the vascularization of the cornea and congestion of the conjunctiva are visible to the naked eye.

Riboflavin is synthesized by plants and occurs particularly in germinating seeds and young growing green plants. Peas, beans, grains and yeast are the most important plant sources. Animal sources of value are milk, egg-yolk, liver, kidney and heart muscle (Table 10, 11): the riboflavin found in most plant and animal tissues, as well as in yeast, is mainly present in the combined form but in milk the greater part is in the free state. Many anaerobic bacteria produce riboflavin in large amounts and it is likely that biosynthesis of riboflavin by the flora of the human intestine provides nutritionally important quantities.

TABLE 10, 11

Riboflavin in Foods. mg. per 100 g.

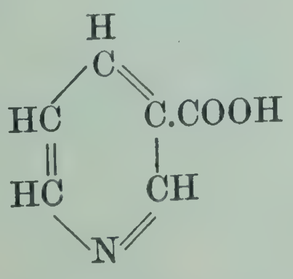
Yeast	5·0	Egg—fresh	0·4
Wheat	0·18 to 0·25	dried	1·3
Barley	0·12 to 0·25	Meat	0·25
Pulses	0·15 to 0·3	Liver	3·0
Milk—whole	0·15 to 0·20	Fruit	0·03
powder	1·15 to 1·60	Vegetables	0·05
Cheese	0·50		

The vitamin is relatively stable in acid solution in the dark. Ordinary cooking processes cause little destruction of riboflavin although a certain amount may be lost by extraction in the cooking water. Considerable loss of riboflavin may occur in milk exposed to bright sunlight.

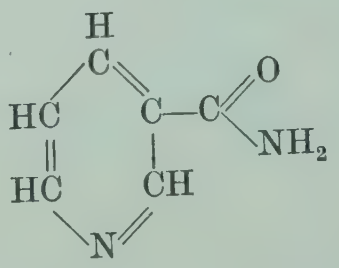
NICOTINIC ACID

Nicotinic acid has been familiar to organic chemists for many years as a degradation product of the alkaloid nicotine but it was not known to possess any nutritional significance until Elvehjem, Madden, Strong and Wooley showed in 1937 that it was capable of curing a deficiency disease in chickens, dogs and man. In 1915 Goldberger had demonstrated that *pellagra* was a deficiency disease but at that time it was believed to be due to a lack of protein or amino-acids. Recent experiments have indeed shown that tryptophan can be converted into nicotinic acid by the rat and that this conversion is accelerated by pyridoxine, thiamine and riboflavin.

The discovery that nicotinic acid cured 'black-tongue' in dogs, a condition developing in that animal on diets which produce pellagra in man, led to its successful use in treatment. The amide of nicotinic acid was isolated from heart muscle by Kuhn and Vetter in 1935 and was later shown to be a component part of coenzyme I (DPN) and of coenzyme II (TPN) (p. 85); it is an essential growth factor for many micro-organisms. Resistant to heat, light, oxidation and alkalis, nicotinic acid is one of the most stable of the vitamins.



nicotinic acid
(β-pyridine-carboxylic acid)



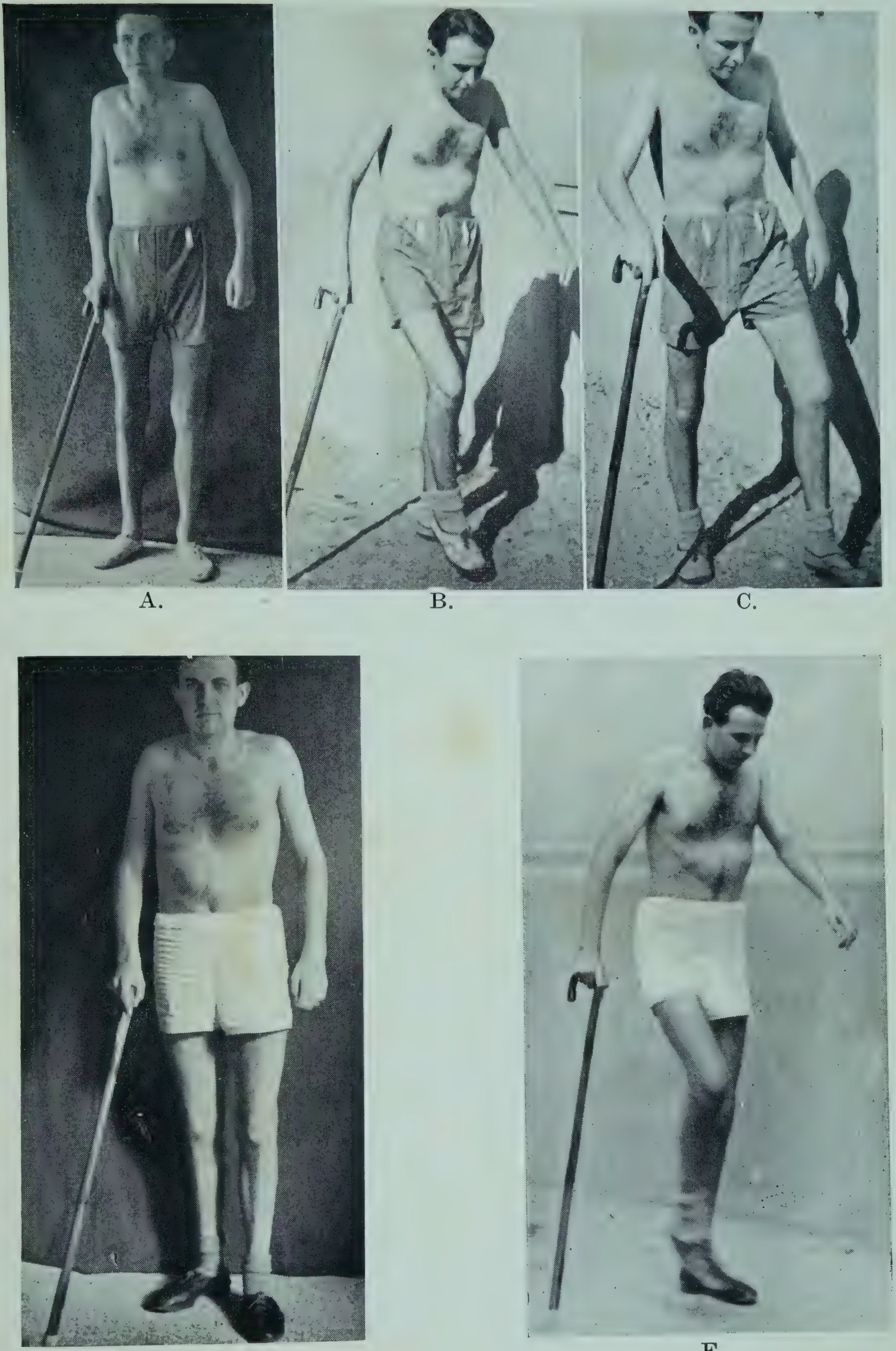
nicotinic acid amide
(nicotinamide)

The beneficial effects produced by the administration of nicotinic acid to a patient or to an animal suffering from deficiency are due to its conversion into the amide which is then converted into coenzyme I and coenzyme II. A number of other pyridine derivatives, such as nicotinic acid diethyl amide (nikethamide), are also effective in the treatment of deficiency disease, presumably because they are converted into nicotinamide in the body.

Both chemical and micro-biological methods are available to determine its concentration in body fluids. Nicotinic acid is present in almost all tissues, the greater proportion being present as coenzyme. The highest concentration is found in the liver. Blood normally contains some 0·30 to 0·83 mg. per 100 ml. mainly as coenzyme, but even when symptoms of deficiency are present the concentration of nicotine acid in the blood is not greatly reduced and appreciable amounts of the vitamin are still excreted in the urine.

Since nicotinic acid has many functions in the body, its deficiency might be expected to show itself in many different ways. However, nicotinic acid deficiency is usually met with in circumstances in which lack of other vitamins, especially those of the B group, is almost certain to occur. It is difficult, therefore, to assign particular features specifically to a deficiency of nicotinic acid since they might be the result of other concomitant deficiencies. However, because it is rapidly relieved by the administration of nicotinic acid or of nicotinamide alone, pellagra must be largely due to a deficiency of this vitamin. Pellagra affects many millions of people every year, being especially prevalent in maize-eating countries where extreme poverty allows only a very inadequate diet. Following a prolonged period of general ill-health, characterized by irritability, depression, loss of weight and strength and lack of appetite, the patient often develops an intractable diarrhœa which accelerates the loss of weight and strength. About the same time the skin becomes reddened and later roughened and scaly. This so-called pellagrous dermatitis is very characteristic, being confined to the exposed parts of the body, face, neck and hands, normally subjected to the action of light and heat (Fig. 10, 15). Subsequently the lesions become brownish in colour and not infrequently infected. Extreme redness and soreness of the mouth, with swelling and redness of the tongue, further increase the patient's difficulties and a vicious circle occurs, in which the more severe the deficiency the more difficult it becomes for the patient to eat a curative diet and the more rapidly therefore the disorder progresses. Mental symptoms such as irritability, anxiety, depression and confusion also occur but they vary very much between one patient and another, and even in the same patient from time to time. In severe cases there is profound mental disorder which, if untreated, may lead to dementia. Most of these disturbances can be corrected within a few days by the administration of nicotinic acid or its amide. This, however, does not mean that the patient is cured. Nearly all pellagrins have lesions of the skin and mucous membranes, such as those already described in riboflavin deficiency, which can be treated only by supplying the appropriate vitamin. As with other deficiency disorders the best treatment for pellagra is a full diet of high vitamin content.

In its combined form as coenzymes I and II nicotinic acid is widely distributed in animal tissues. Liver, kidney and muscle are the best animal sources. Yeast and the outer parts of grain are good natural sources of the vitamin, but the valuable part of the grain is almost com-



A. B. C.

D. E.

Fig. 10, 12

German prisoner of war with nutritional neuropathy. A. shows appearance on 10/9/44, unable to stand unsupported (and grossly ataxic). B. and C. show appearance on 20/10/44, after one month's treatment; general condition approved, gait ataxic. D. and E. show condition on 17/1/45; further general improvement and gait better. (J. D. Spillane (1947), *Nutritional Disorders of the Nervous System*, p. 161. Edinburgh: Livingstone.)

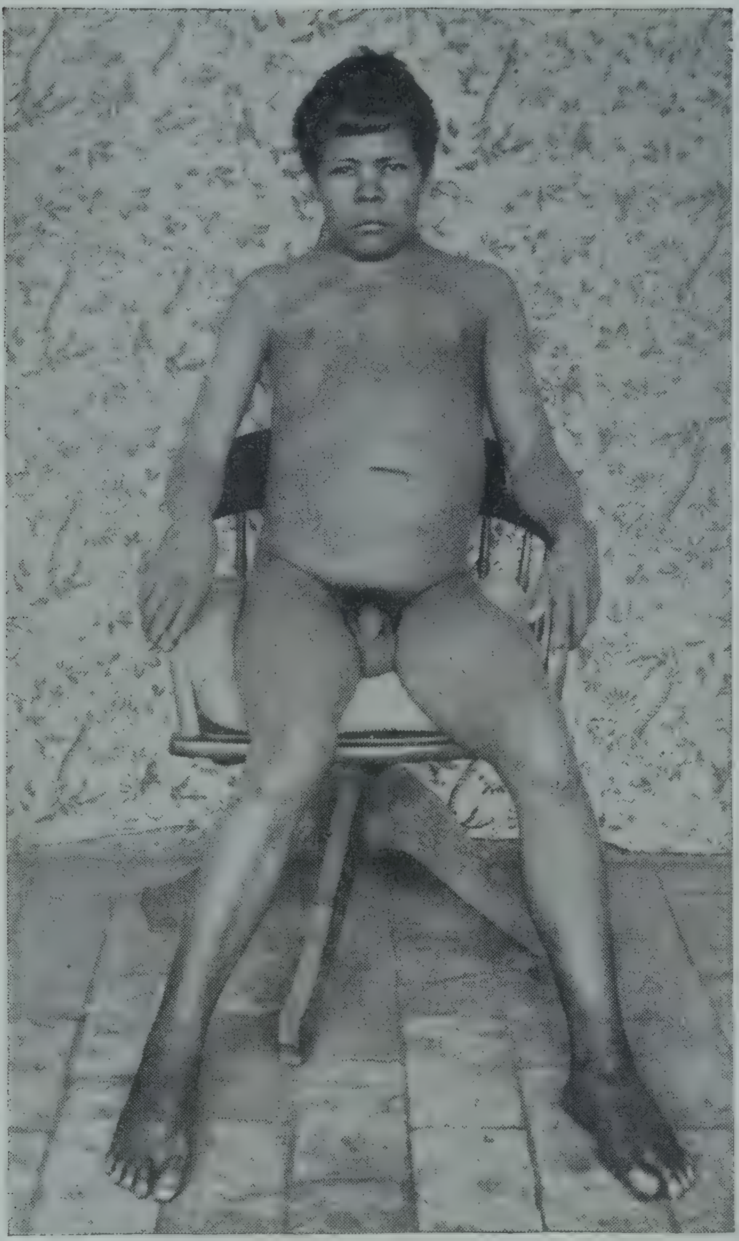


FIG. 10, 13

A case of beri-beri with extensive œdema. (*Courtesy of Wellcome Museum of Medical Science.*)

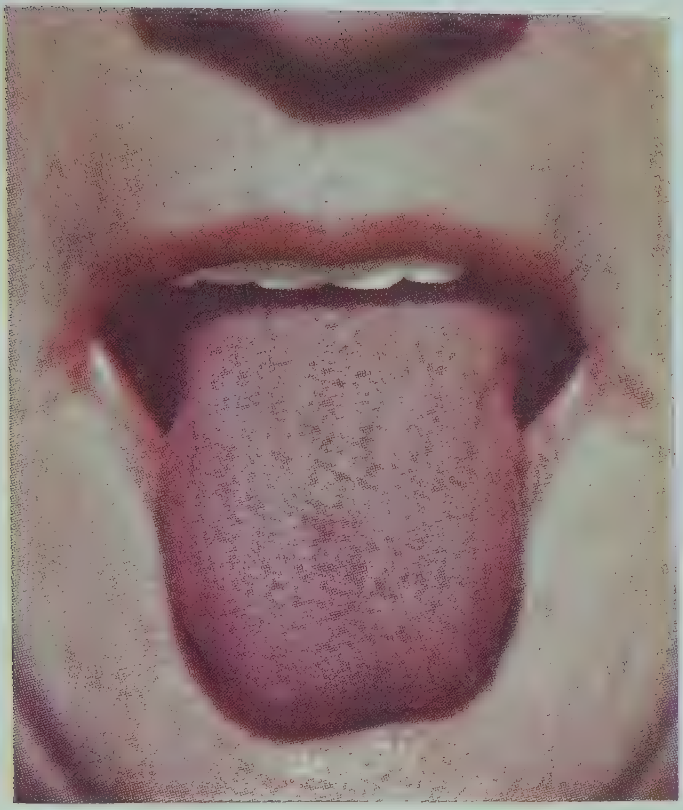


Fig. 10, 14A

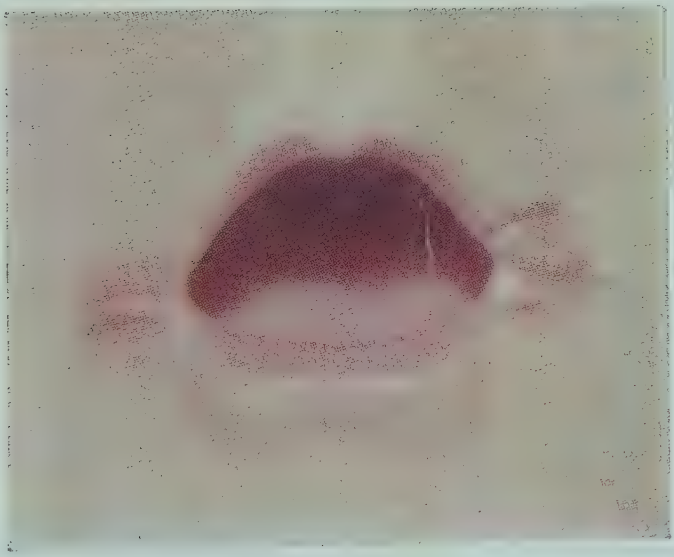


Fig. 10, 14B

Typical lesions of the riboflavin deficiency syndrome; 14A. Tender magenta tongue; 14B. Angular stomatitis. ("*Roche Courier*" (1951), **17,** 5 *and by courtesy of the Department of Surgery, The University of Liverpool.*)

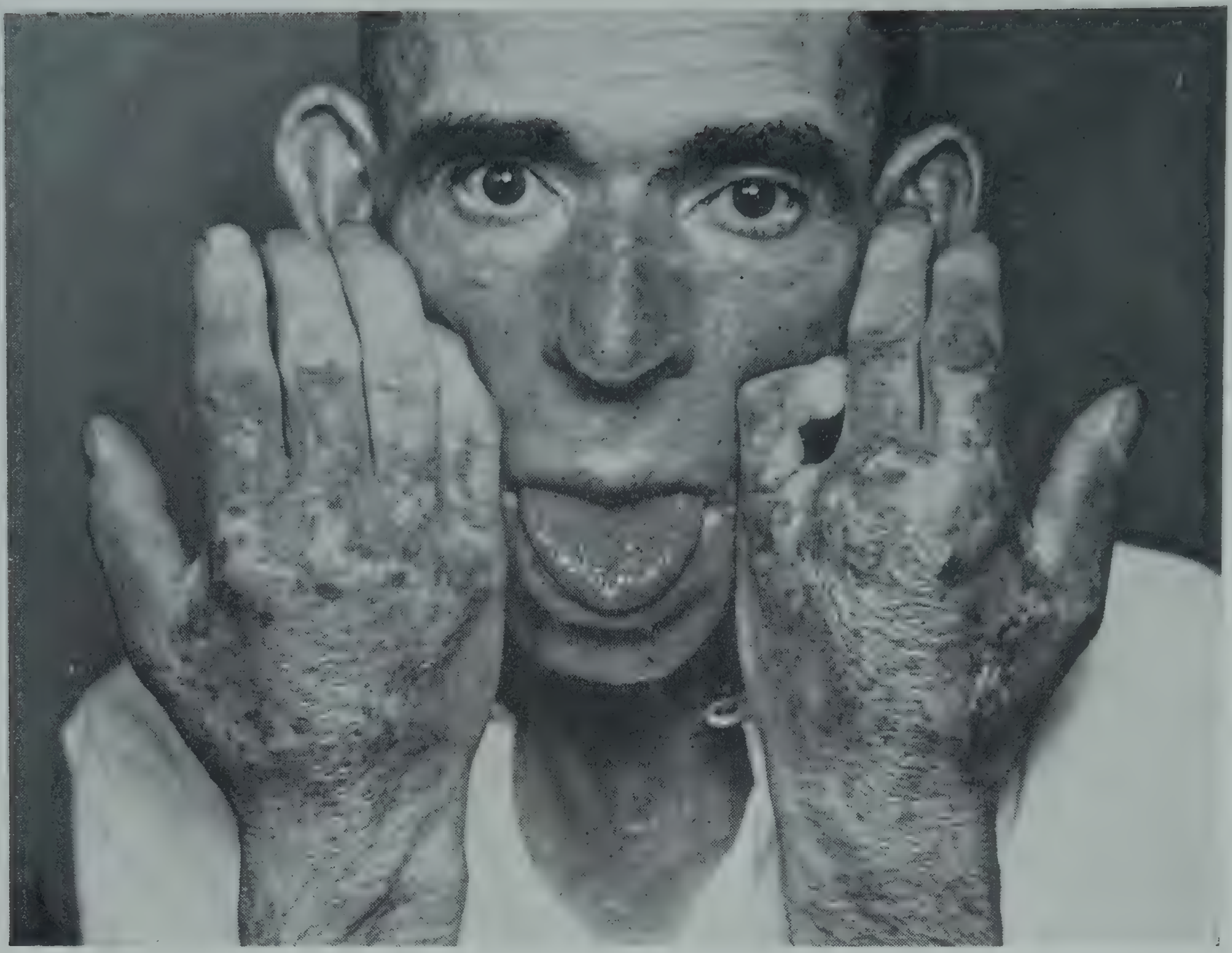


FIG. 10, 15A

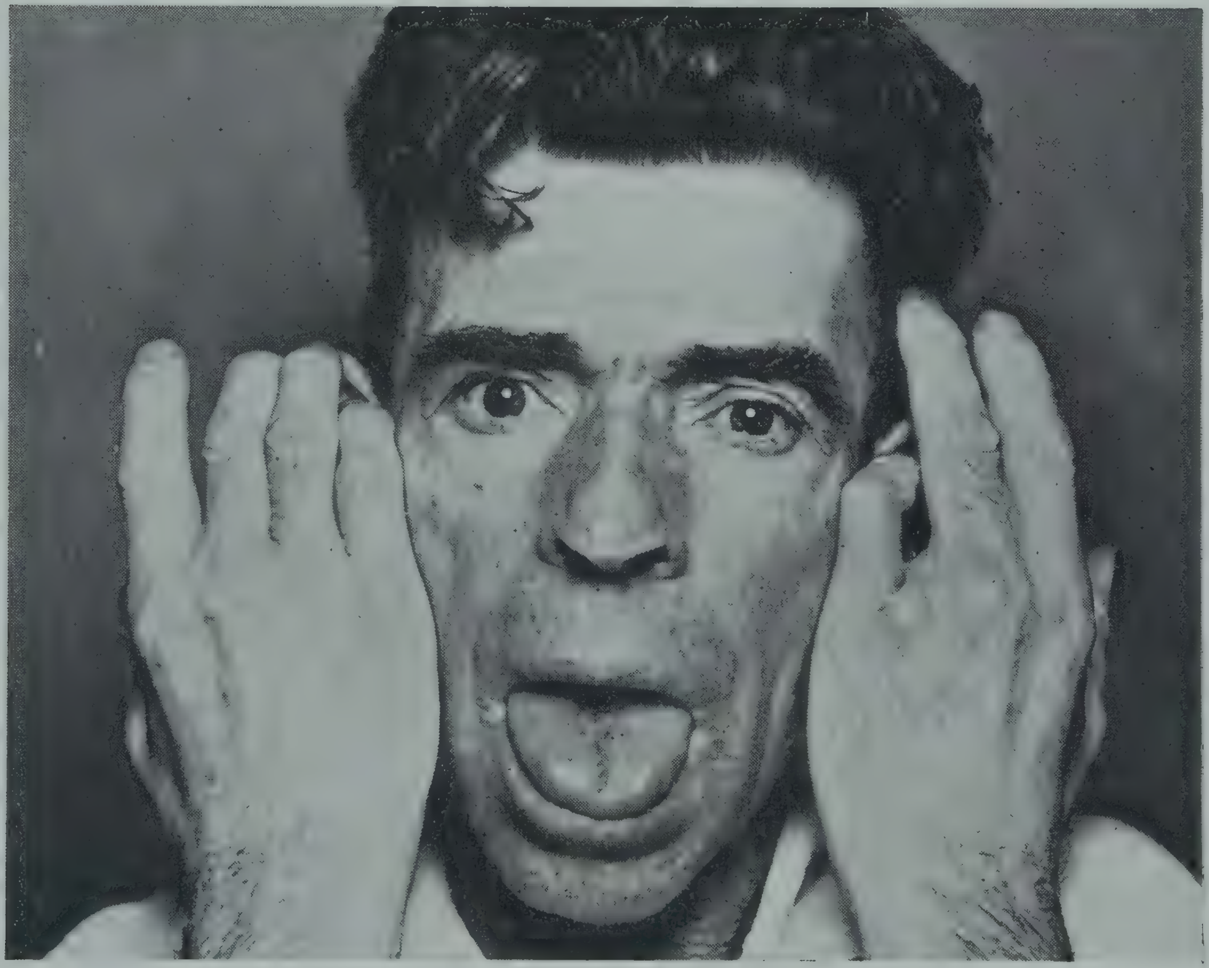


FIG. 10, 15B

Showing the dermatitis of face and hands in a case of pellagra subsequently cured by administration of nicotinic acid amide. Note the difference in the appearance of the tongue, which before treatment was red and tender (R. W. Vines and A. M. Olsen (1943), *Proc. Staff Meetings Mayo Clinic*, **18**, 389).

pletely removed as bran in the milling process (Table 10, 16). Like riboflavin, nicotinic acid is synthesized by intestinal bacteria. Since

TABLE 10, 16

Nicotinic Acid in Foods. mg. per 100 g.

Yeast, dry . . .	10 to 50	Kidney . . .	6 to 10
Wheat, whole . .	2·8 to 5·0	Fresh meat . . .	3 to 10
Flour, white . . .	0·6 to 0·8	Cow's milk . . .	0·1 to 0·5
85 per cent. extraction	0·15 to 0·20	Bread—whole grain .	5·0
Bran	30 to 35	white . .	1
Liver, fresh . . .	10 to 27	Fish	3 to 8

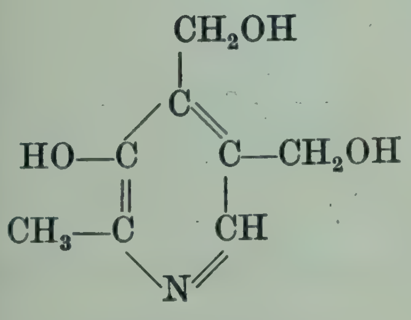
nicotinic acid is one of the most stable of the vitamins, losses in the preparation and cooking of food are small. Significant losses may, however, occur by extraction into the cooking water.

PYRIDOXINE, PANTOTHENIC ACID AND BIOTIN

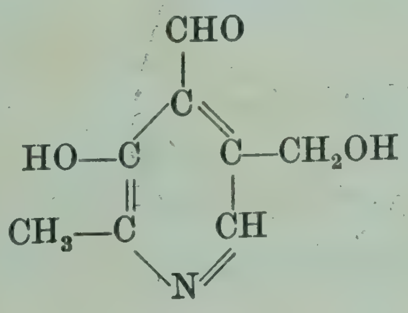
These members of the B complex, though very different in chemical and physical properties, may be considered together because deficiency of each of them has been identified either with a lesion of the skin or with disturbances of the pigmentation of the skin and hair. With the exception of biotin they have not been assigned any definite role in human nutrition.

All three vitamins are found widely distributed in both animal and vegetable foods, more particularly in those rich in other members of the B group. Yeast, liver, kidney, peas, and the outer coats of grains are the richest natural sources of the group.

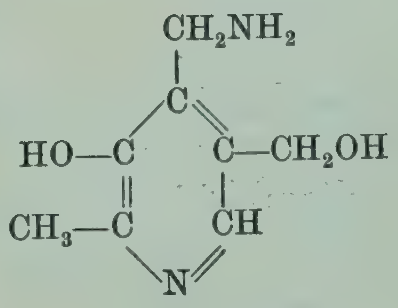
Pyridoxine (adermin, vitamin B_6), which is a primary alcohol, is found in nature in association with the corresponding aldehyde (pyridoxal) and amine (pyridoxamine). Like nicotinamide these substances are all derivatives of pyridine.



pyridoxine



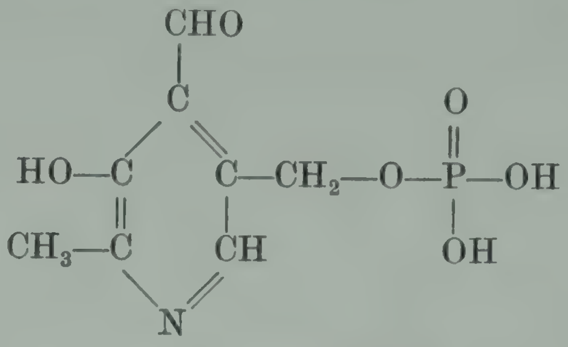
pyridoxal



pyridoxamine

These three compounds have similar biological properties and are necessary for the growth of certain bacteria, moulds and yeasts. It is probable that pyridoxine and pyridoxamine are converted to pyridoxal in the tissues. Just as riboflavin and thiamine can be phosphorylated to yield coenzymes, so pyridoxal yields pyridoxal phosphate which acts as coenzyme for some amino-acid decarboxylases (p. 350) and for transaminase (p. 341).

Pyridoxal phosphate catalyses the synthesis of tryptophan from indole and serine and, in the rat, is necessary for the conversion of tryptophan to nicotinic acid. Both chemical and micro-biological methods are available for its estimation.

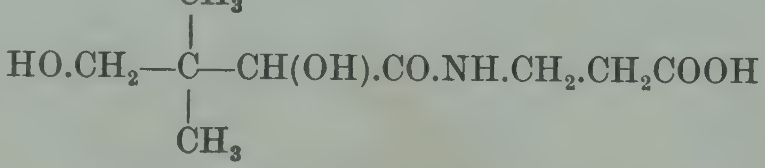


pyridoxal phosphate

Pyridoxine has been shown to be an essential nutrient for the rat, chicken, pigeon, dog and pig. The most characteristic deficiency lesions occur in rats which develop soft, reddened, swollen areas in the paws and mouth from which the thin matted fur finally disappears completely (*acrodynia*). Ulceration of the mouth and tongue, thickening and swelling of the ears and extreme emaciation are other features of the syndrome, which is rapidly cured by the administration of the vitamin. Pyridoxine deficiency in dogs and in pigs results in severe anæmia and in deficient pigs and rats epileptiform fits have occurred. No clear-cut symptoms resulting from pyridoxine deficiency have yet been described in man although symptoms, relieved at once by the administration of pyridoxine, have been produced by giving the anti-vitamin (p. 134) deoxypyridoxine.

On the basis of animal experiments it has been suggested that the daily requirements for man are about 1·5 to 2 mg.

Pantothenic Acid, so called because of its widespread distribution, was isolated in 1939 and was identified as α, γ-dihydroxy-β, β-dimethylbutyryl-β-alanine:



$$\mathrm{HO.CH_2{-}\overset{\displaystyle CH_3}{\underset{\displaystyle CH_3}{C}}{-}CH(OH).CO.NH.CH_2.CH_2COOH}$$

pantothenic acid

Pantothenic acid acts in the living cell as a component of the important substance, coenzyme A (p. 62), which plays an essential role in the metabolism of both fat and carbohydrate (Fig. 20, 19).

Liver, kidney, meat, wheat bran, peas and yeast are good sources of pantothenic acid.

Pantothenic acid is an essential nutrient for rats, mice, chicks, dogs and pigs, as well as for yeast and certain micro-organisms (notably streptococci). Chickens show characteristic lesions in pantothenic acid deficiency, namely a specific dermatitis associated with widespread degeneration in the spinal cord. Lack of the vitamin in the diet of rats results in marked loss of weight, symmetrical greying of the fur (*achromotrichia*) (Fig. 10, 17), loss of fur and irritation of the nose and eyes, together with hæmorrhage into, and necrosis of, the adrenal glands. Similar hæmorrhagic lesions in the

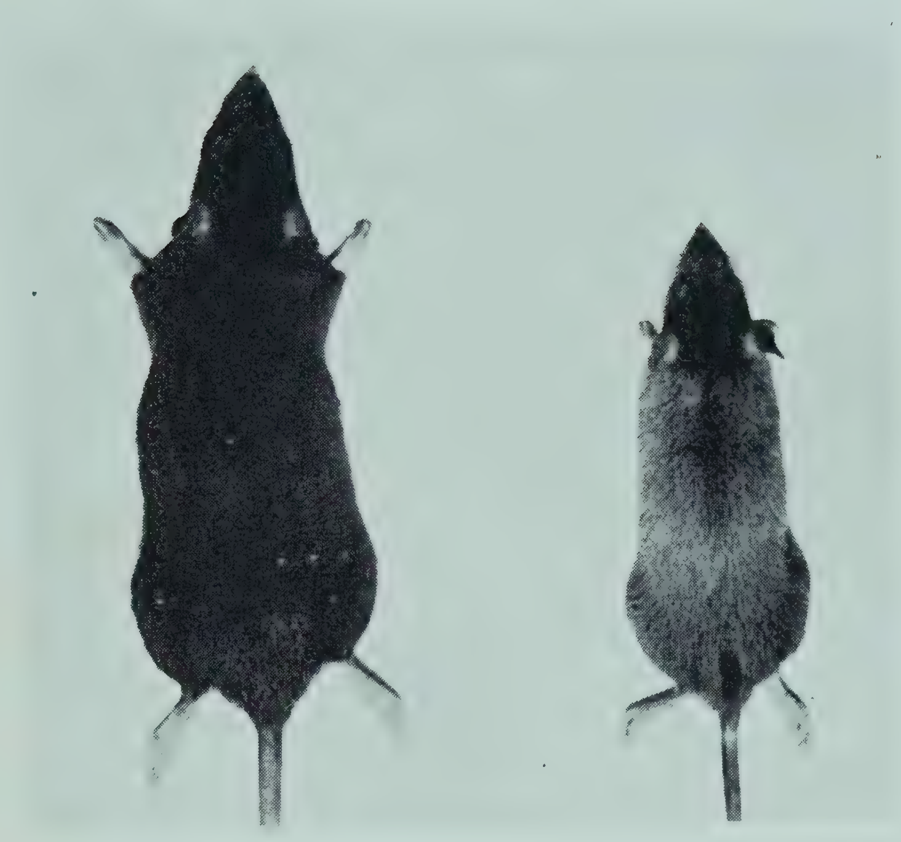


Fig. 10, 17

The two rats shown in the photograph were fed for 35 days on a diet deficient in B complex but supplemented with nicotinamide, thiamin, riboflavin and vitamin B_6 The animal on the left received a daily dose of 100 micrograms of synthetic dextrorotatory calcium pantothenate as well. Both animals were of the same size when the experiment started. Nutritional achromotrichia may be due to other factors as well as pantothenic acid but it is obvious that pantothenic acid has prevented greying in this instance. (From Eddy, W. H. & Dalldorf, G. (1944), *The Avitaminoses*, 3rd edition. Baltimore: Williams & Wilkins Co.)

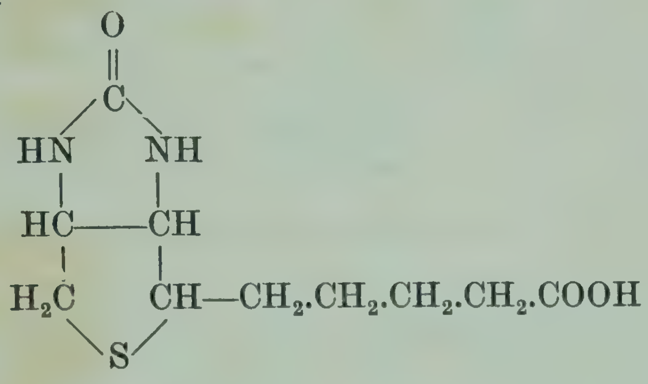
(*By courtesy of Dr. Klaus Unna.*)

idneys and alimentary tract, together with fatty degeneration in the liver, ave been reported in deficient dogs.

There is no evidence that pantothenic acid or any other nutritional actor can influence either the greying or the loss of hair in man.

Biotin, a relatively stable cyclic ureide, was isolated from egg-yolk n 1936 and proved in 1940 to be identical with vitamin H, the protective actor against egg-white injury extensively studied by György.

Biotin is essential for the growth of many micro-organisms, moulds nd yeasts; it is a growth factor for plants and for many animals and it rotects against egg-white injury in rats, chicks, dogs, monkeys and man.



biotin

The mode of action of biotin is not so clearly established as that of other enzymes but it is believed to act as a coenzyme in relation to the fixation of carbon dioxide in such reactions as the carboxylation of pyruvic acid to oxaloacetic acid.

$$CO_2 + CH_3COCOOH \rightleftharpoons HOOCCH_2COCOOH$$

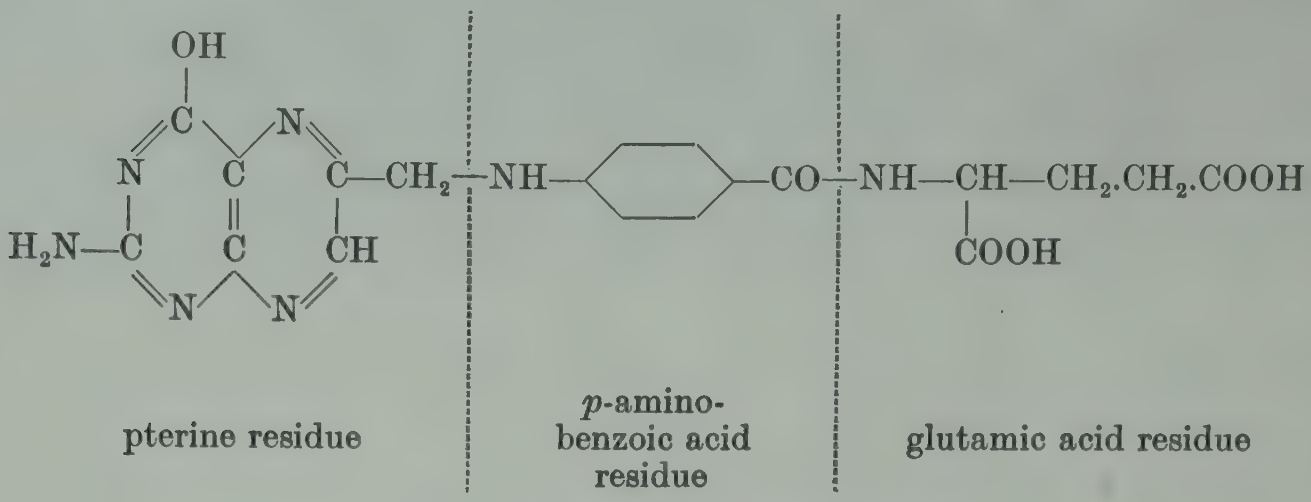
It is also required in the enzymic deamination of some amino-acids.

It has been known since 1916 that a high concentration of uncooked egg-white in experimental diets is toxic to rats. The reason is that the protein avidin of egg-white forms with biotin a stable compound which, since it is not hydrolysed in the intestine, cannot be absorbed. Thus the real basis of egg-white injury is biotin deficiency, which in the rat is characterized by a generalized dermatitis, great loss of weight and a progressive spastic paralysis. Biotin deficiency has been produced in human volunteers by a diet poor in biotin but containing large amounts of raw or dried egg-white. A case has been described in a man whose diet consisted mainly of six dozen raw eggs a week and four quarts of wine daily. He had a severe generalized dermatitis with loss of hair and sebaceous secretion; these disturbances disappeared soon after he was given injections of biotin methyl ester. It is clear, therefore, that biotin is an essential food factor in man but biotin deficiency is likely to occur in only the most exceptional circumstances.

FOLIC ACID, VITAMIN B_{12} (COBALAMINS), AND *P*-AMINOBENZOIC ACID

Folic Acid.—The name folic acid is applied to a factor important for the growth of certain bacteria which was first obtained from green leaves. It has also been shown to be a vitamin essential for the growth of chicks and to be an anti-anæmia factor for chicks, monkeys and man.

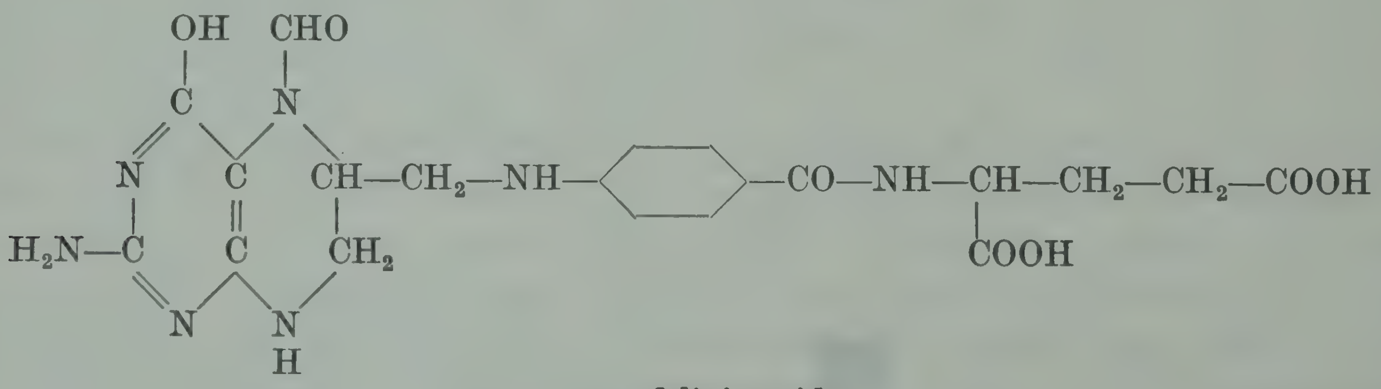
Included in the term 'folic acid' is a group of chemically distinct substances. The simplest of these is pteroyl-glutamic acid (PGA), which has been synthesized, but folic acid can also exist in the form of conjugates of pteroyl-glutamic acid, in which there are several glutamic acid residues.



folic acid (pteroyl-glutamic acid)

Folic acid is synthesized in nutritionally significant amounts by bacteria inhabiting the human intestine and is widely distributed in foodstuffs in conjugated form. Pteroyl-glutamic acid can produce partial or even complete remission of pernicious anæmia and certain other macrocytic anæmias in man, but it is not the anti-pernicious anæmia factor of liver, since highly active concentrates of liver may contain almost no folic acid. Folic acid is unable to cure or prevent the degeneration of the spinal cord that is found in association with pernicious anæmia.

There is reason to believe that folic acid is converted in the living cell into *folinic acid*, also known as *leucovorin* or *citrovorum factor* (*CF*) which has been synthesized. It has the following structure but other similar forms may also exist such as N^{10}-formyl PGA.



folinic acid
(5-formyl-5, 6, 7, 8-tetrahydrofolic acid)

Folic acid may be converted into folinic acid under the influence of enzymes present in liver and kidney. Ascorbic acid is essential for this process.

By virtue of its formyl (—CHO) grouping folinic acid is implicated, in conjunction with vitamin B_{12}, in the metabolism of the one-carbon fragments involved (*a*) in the biosynthesis of purines (carbon atom numbers 2 and 8 in the purine ring), (*b*) in the methylation of ethanolamine to give choline or of the pyrimidine ring to give thymine, (*c*) in the formation of serine from glycine and (*d*) in the biosynthesis of histidine.

Vitamin B_{12} (Cyanocobalamin).—Many attempts have been made to isolate the antipernicious anæmia factor (p. 394) from liver, but although ighly purified preparations have been available for a long time it was not ntil 1948 that the factor was finally obtained from liver (20 mg. from one on) in the form of a pure, red-coloured, crystalline substance of molecular

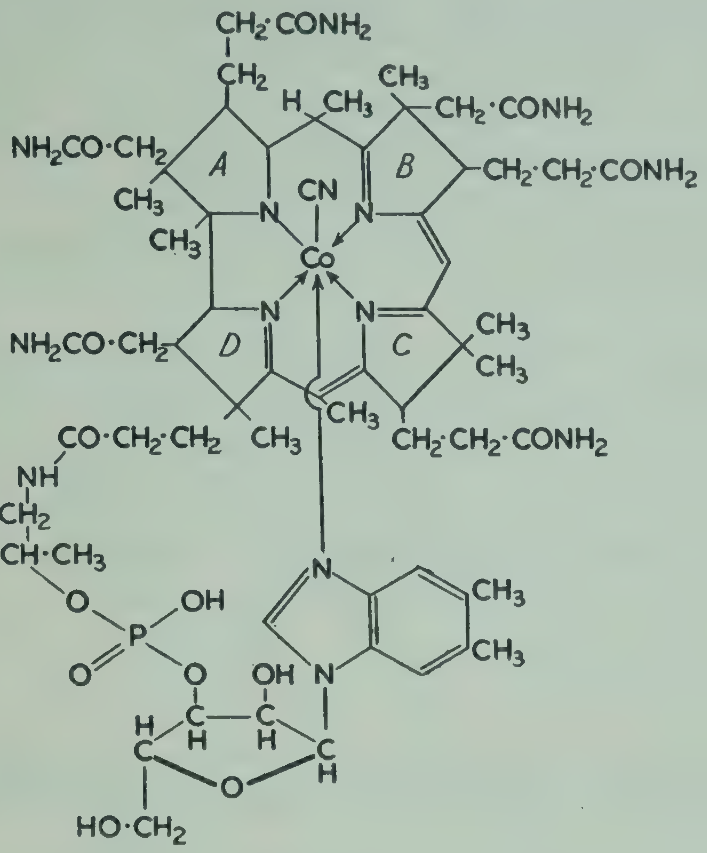

CH₂·CONH₂
CH₂
H
CH₃
CH₃
CH₂·CONH₂
NH₂CO·CH₂
A
B
CH₂·CH₂·CONH₂
CH₃
CN
CH₃
N
N
Co
N
N
CH₃
NH₂CO·CH₂
D
C
CH₃
CH₃
CO·CH₂·CH₂
CH₃
CH₂·CH₂·CONH₂
NH
CH₂
CH·CH₃
O
OH
P
O
O
OH
N
N
CH₃
CH₃
H
C
H
C
H
C
H
C
H
O
HO·CH₂

Fig. 10, 18

Vitamin B_{12} (cyanocobalamin)

weight about 1350 containing some 4 per cent. of cobalt. The molecule of vitamin B_{12} contains the base dimethyl-benziminazole linked to the pentose sugar ribose and to phosphoric acid to form a nucleotide structure. This in turn is linked to a porphyrin ring system (p. 411) containing one atom of cobalt (Fig. 10, 18).

Table 10, 19

Vitamin B_{12} in foods, μg. per 100 g.

Various fresh meats . .	1 to 3
Beef, kidney and liver . .	15 to 20
Egg yolk	1·4
Cheese	1·4
Dried casein	1·0
Rumen contents . . .	50
Milk	traces

Vitamin B_{12} is effective in curing both the hæmatological and neurological features of pernicious anæmia in doses as small as 1 μg. per day. It also acts as a growth-promoting factor for a micro-organism known as *Lactobacillus lactis* (Dorner) which can therefore be used for assay purposes in place of the tedious clinical tests which previously had to be employed. Vitamin B_{12} is synthesized by a micro-organism known as *Streptomyces griseus*, and the various factors, B_{12}, B_{12b}, B_{12c}, have been obtained from *S. griseus* fermentation liquors by partition chromatography. This group of compounds is now known collectively as the cobalamins. B_{12} is cyanocobalamin, B_{12b} is hydroxocobalamin, and B_{12c} is nitritocobalamin.

The role of this vitamin and of folic acid in the maturation of blood cells, as well as the relationship between vitamin B_{12}, folic acid, and the intrinsic and extrinsic factors of Castle (p. 394) have not yet been fully worked out (p. 395). It is known, however, that vitamin B_{12}, probably in conjunction with folic acid, is intimately associated with the synthesis of methyl groups and with methyl group transfer.

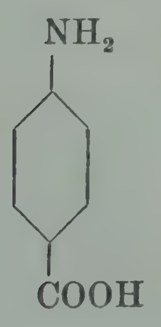
The amounts of vitamin B_{12} required to maintain health in man are not known.

p-Aminobenzoic acid forms the substrate of the enzyme system required for the synthesis of folic acid. It is a growth factor for the chick and for many species of bacteria. Diets lacking *p*-aminobenzoic acid produce greying of the fur in black or piebald rats. Although for these reasons *p*-aminobenzoic acid is usually classified as a vitamin it has no known role in human nutrition.

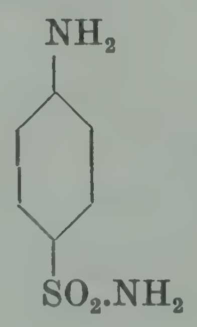
COMPETITIVE INHIBITION AND ANTI-VITAMINS

We have already seen (p. 75) that the enzymic oxidation of succinic acid is inhibited by malonic acid. The reason for this inhibition, referred to as *competitive inhibition*, is that the molecule of malonic acid resembles the succinic acid molecule so closely that it becomes adsorbed to the surface of the enzyme, succinic dehydrogenase, so that adsorption of succinic molecules is blocked and the oxidation of succinic acid thereby retarded or even prevented.

The phenomenon of competitive inhibition, or *substrate competition* as it is also called, has a wider application in biology and clinical medicine. For example, the effectiveness as antibacterial agents of the sulphonamide group of drugs, of which *p*-aminobenzene sulphonamide is a simple example, is now known to be due to their chemical similarity to *p*-aminobenzoic acid which, as we have seen, is indispensable to the growth of many bacteria.

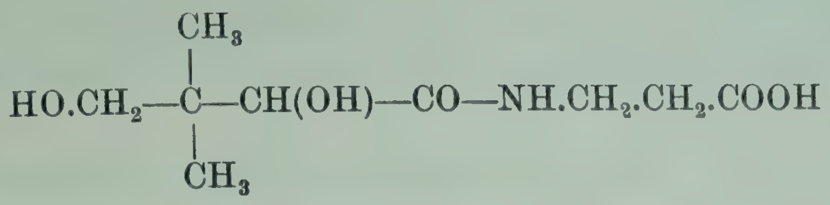


p-aminobenzoic acid



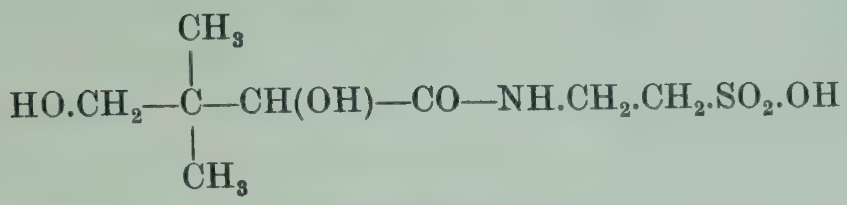
p-aminobenzene sulphonamide
(sulphanilamide)

The sulphonamide molecule becomes incorporated into an enzyme system which is thereby blocked so that the metabolism and multiplication of the bacteria are interfered with. In order to exert this blocking effect the sulphonamide must be present in concentrations 1500 to 2000 times as great as those of *p*-aminobenzoic acid. It follows, of course, that the bacteriostatic properties of sulphonamides are annulled by high concentrations of *p*-aminobenzoic acid. By a similar mechanism the growth of those organisms (hæmolytic streptococci and pneumococci) which require pantothenic acid as an essential metabolite is inhibited by pantoyl taurine.



$$\mathrm{HO.CH_2{-}\overset{\displaystyle CH_3}{\underset{\displaystyle CH_3}{C}}{-}CH(OH){-}CO{-}NH.CH_2.CH_2.COOH}$$

pantothenic acid



$$\mathrm{HO.CH_2{-}\overset{\displaystyle CH_3}{\underset{\displaystyle CH_3}{C}}{-}CH(OH){-}CO{-}NH.CH_2.CH_2.SO_2.OH}$$

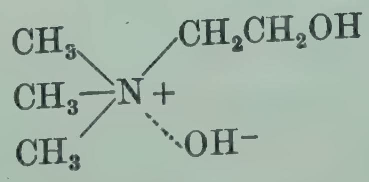
pantoyltaurine

The important biochemical action of thiamine (p. 122) can be blocked by pyrithiamine, the analogue of thiamine in which the pyrimidine ring system is replaced by pyridine. This compound may therefore be described as an 'anti-vitamin'. Substances antagonizing the action of thiamine have been found in cereals. An early observation that pellagra tended to occur in areas where maize formed a high proportion of the diet has recently been confirmed experimentally in animals. It is now suggested that the 'toxic' substance, or anti-vitamin, in maize is an analogue of nicotinic acid which is incorporated, like the vitamin, into the coenzyme molecules, but which renders them incapable of acting as hydrogen acceptors.

LIPOTROPIC FACTORS

The deposition of an excessive amount of fat in the liver of an experimental animal or of man may be produced by feeding diets containing a very high proportion of lipid, or by poisons such as phosphorus. In rats the fatty infiltration occurs much more readily on diets deficient in the vitamins of the B complex and can be prevented by the addition of choline or the amino-acid methionine, which are therefore said to exert a *lipotropic* action. The exact mechanism of the lipotropic action of choline, which is shared also by inositol, is unknown, but it is believed that the lipotropic factors promote the transport of fatty acids as phospholipids from the liver to the tissues or accelerate the utilization of fat in the liver itself.

Choline is essential for the normal nutrition of the chicken and rat as well as for the rabbit and the dog. It is used in the synthesis of acetylcholine and phospholipids.

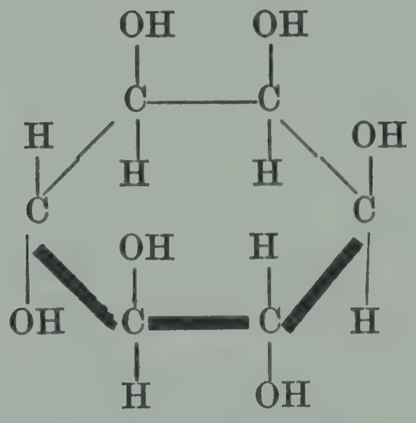


choline

The methyl groups of choline play an important role in intermediary metabolism since they are concerned in the process of transmethylation by which methyl groups are transferred from one metabolite to another (p. 347). Provided some other source of utilizable methyl group is available it is not clear that choline is an essential part of the diet of man. In any case, choline deficiency is extremely unlikely to occur because lecithin, which contains choline, is widely distributed in all animal and vegetable foods.

Indeed it is doubtful whether choline should be regarded as a vitamin at all. Although it is admittedly essential for nutrition it is required in comparatively large amounts (50 to 100 mg. per kg. body weight) and is, in the form of lecithin, part of the biochemical structure of the body.

Inositol is present in plants as phytic acid (inositol hexaphosphate) which is mentioned on p. 13. It can exist in several stereoisomeric forms but the one which is biologically active is *meso*-inositol.



meso-inositol

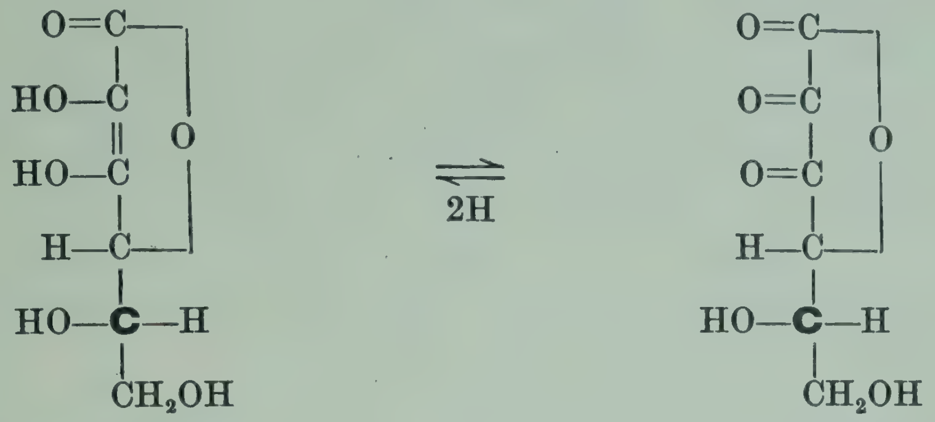
Inositol is widely distributed both in animal (liver, kidney, meat, eggs) and vegetable foods (fruit, cereals); it is present abundantly also in yeast and in many moulds and bacteria. It occurs in combined form in certain phospholipids (lipositols, p. 23). For this reason a human deficiency syndrome is unlikely and has not been described. Inositol is, however, probably essential to the rat and mouse for the growth of hair and for proper nutrition of the skin.

The insecticide lindane or gammexane is similar in structure to inositol. It may therefore produce its toxic effect by acting as an anti-metabolite.

ASCORBIC ACID (VITAMIN C)

The first controlled experiment from which our present knowledge of ascorbic acid derives was made in 1753 by Lind who showed that scurvy (scorbutus), which was in those days so common and so severe a disease among seafarers, could be cured by giving the patients oranges and lemons. During his voyages round the world between 1772 and 1775 Captain Cook kept his ship's company free from scurvy by giving them abundant fresh food. Scurvy was abolished from the British Navy in 1804 as the result of the use of lime juice on long voyages, hence the name 'limey' which is still used abroad to describe a British seaman. When Holst and Frölich in 1907 produced scurvy experimentally in guinea-pigs it became possible to study the antiscorbutic vitamin in more detail.

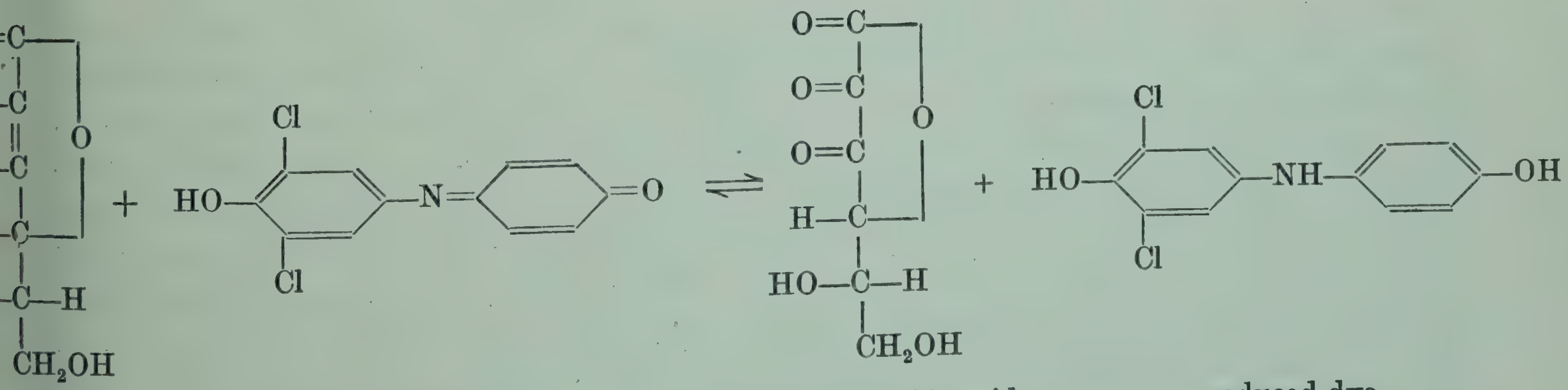
In 1932 ascorbic acid was isolated from lemons by Waugh and King and was found to be identical with a compound obtained from adrenal glands and from cabbage in 1928 by Szent-Györgyi, but not recognized by him as the vitamin. The structural formula of ascorbic acid was worked out by Haworth and Hirst in 1933 and the vitamin was synthesized in the same year.



L-ascorbic acid L-dehydroascorbic acid

Ascorbic acid possesses an asymmetric carbon atom **C** and is therefore optically active. The vitamin is the L-isomer, the D-compound being biologically inert. Although compounds very similar in structure have been synthesized they are all less active biologically and L-ascorbic acid is the only known naturally occurring antiscorbutic substance.

Pure ascorbic acid is a white crystalline solid freely soluble in water. It has a pleasant acid taste. It is a powerful reducing agent (it reduces Fehling's solution in the cold), giving up two hydrogen atoms to become dehydroascorbic acid. This oxidation is reversible in the body so that L-dehydroascorbic acid is fully antiscorbutic. Ascorbic acid, is, however, easily oxidized beyond the stage of dehydroascorbic acid to give products (diketogulonic and oxalic acids) from which the vitamin cannot be recovered by reduction in the body. The usual chemical methods for the determination of ascorbic acid in biological material are based upon its powerful reducing properties. Thus under suitable conditions the blue dye 2 : 6-dichlorophenol-indophenol can be reduced to a colourless compound by ascorbic acid. The reaction is not however specific to ascorbic acid.



bic acid 2 : 6-dichlorophenol-indophenol (blue) dehydroascorbic acid reduced dye (colourless)

The presence and position of ascorbic acid in a tissue, or even in a single cell, can be detected by submitting the preparation to a weak solution of silver nitrate. Parts containing the vitamin are indicated by deposits of silver.

The stability of ascorbic acid in solution, or in foodstuffs, is affected by the pH, the temperature, contact with oxygen and traces of metals, especially copper. In the presence of oxygen and traces of copper, ascorbic acid is rapidly oxidized at pH values above 4, especially if the solution is warmed. Oxidation is accelerated by ultraviolet light and the presence of riboflavin. Ascorbic acid is thus liable to be destroyed during the preparation and cooking of many of the foods that contain it but the loss of ascorbic acid by leaching into the cooking water is as a rule far greater than the loss by oxidation. Fruit and vegetables contain an enzyme, ascorbic-acid oxidase, which accelerates the oxidation of ascorbic acid in the presence of oxygen. If the foodstuff is bruised or chopped up, this enzyme is liberated from the damaged cells and rapidly oxidizes much of the vitamin originally present.

Ascorbic acid is synthesized by certain moulds, fungi, all the higher plants, and by animals except the guinea-pig, primates and man. The latter are therefore capable of developing deficiency symptoms (scurvy) if the diet is lacking in the vitamin. Ascorbic acid is stored in the body and a proportion of each dose administered to a deficient subject is therefore retained (Fig. 10, 3). Thus the urinary excretion of the vitamin is governed not only by its rate of absorption from the alimentary canal but also by the rate at which it is taken up from the blood into the tissues. When the body is saturated, 50 to 60 per cent. of ascorbic acid ingested is metabolized to dehydroascorbic acid and other oxidation products. In the guinea-pig, too, it has been found by giving ascorbic acid containing radioactive C that 66 per cent. of an ingested dose is broken down, the carbon appearing in the expired air, and 22 per cent. excreted into the urine. In man ascorbic acid appears in the urine when the concentration in the plasma reaches approximately 1 mg. per 100 ml.

Ascorbic acid is present in all body fluids and tissues. In the cells of the adrenal cortex and medulla the vitamin is found aggregated near to the Golgi apparatus. A high content of ascorbic acid is characteristic especially of glandular tissues, thus the suprarenal glands contain some 100 to 200 mg. per 100 g. and the pituitary, corpus luteum and thymus are all rich sources; liver, lung and heart muscle possess successively lower concentrations and skeletal muscle has about 2 mg. per 100 g. Of particular interest is the concentration of ascorbic acid in the white blood cells, because it forms a reliable guide to an individual's nutritional status. There is normally some 15 to 30 mg. of ascorbic acid per 100 g. of leucocytes in the blood of healthy persons taking diets rich in the vitamin but in subjects depleted of ascorbic acid the vitamin C content of the leucocytes falls progressively, to reach very low values three to six weeks before the signs of scurvy appear (Fig. 10, 20). The concentration of ascorbic acid normally found in blood plasma varies from 0·5 to 2 mg. per 100 ml., but figures below 0·5 mg. per 100 ml. may be found over a long time in any individual without apparently any sign of ill-health.

Ascorbic acid belongs to a group of compounds which form reversible oxidation-reduction systems and the vitamin has therefore been supposed to act as a hydrogen carrier or respiratory catalyst. There is evidence that it may act in this capacity along with glutathione.

Ascorbic acid *in vitro* inhibits the formation of pigment from adrenaline and from 3 : 4-dihydroxyphenylalanine ('dopa'). It prevents, too, the darkening of cut fruit and potatoes which is the result of the oxidation of phenols to catechols and quinones. There is evidence that in ascorbic acid deficiency the end-product of tyrosine oxidation is *p*-hydroxyphenyl pyruvic acid, and ascorbic acid has been found necessary for the oxidation of this substance by liver tissue. Premature human infants fed on cow's milk and full-term infants fed tyrosine or phenylalanine as supplements excrete *p*-hydroxyphenyl pyruvic acid and *p*-hydroxyphenyl lactic acid unless they are also given ascorbic acid. These facts suggest that ascorbic acid has a role in the metabolism of tyrosine and phenylalanine which has in fact been found to be abnormal in human scurvy.

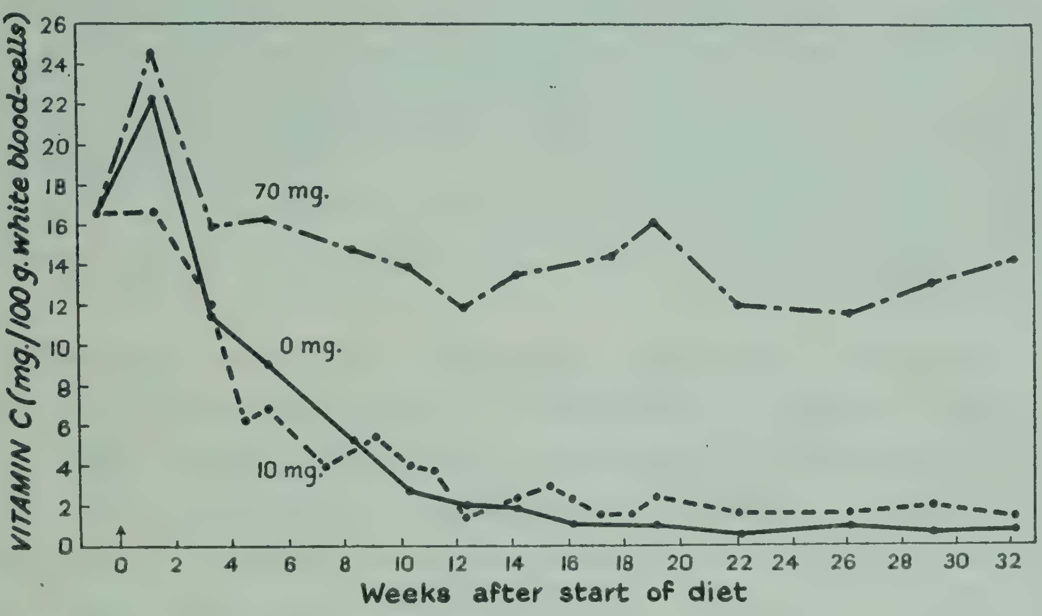


FIG. 10, 20

Mean ascorbic acid concentration in white blood cells, estimated with dichlorophenolindophenol, of groups of volunteers receiving daily 70, 10, or 0 mg. of ascorbic acid as supplements to the basal diet, which contained about 1 mg. per day. In the third group follicular hyperkeratosis appeared after 17–21 weeks and frank bleeding after 26–34 weeks; swelling of, and bleeding from, the gums occurred after 30–38 weeks. No scorbutic signs were observed in the volunteers receiving 70 or 10 mg. ascorbic acid per day (Vitamin C requirement of human adults. A preliminary report by the vitamin C sub-committee of the accessory food factor committee, M.R.C., *Lancet*, (1948), i, 853).

One of the first signs of a deficiency of ascorbic acid is an impaired formation of connective tissue. Connective tissue normally consists of fibroblasts lying in a matrix of collagen. In ascorbic acid deficiency the formation of collagen or intercellular substance is impaired, just as it is also in connective tissue grown in culture. This primary effect of ascorbic acid lack has a number of results.

1. *Bone Formation and Repair*.—In young growing bones, further bone formation at the epi-diaphyseal junctions ceases and the existing bone becomes less dense. Morphological changes occur in the osteoblasts which become functionless. Many revert to fibroblasts and there is increased formation of poorly developed fibrous tissue uniting epiphysis and diaphysis. Under the periosteum, however, continued proliferation of osteoblasts leads to raising of the periosteum and the laying down of cortical bone less dense than normal. In adults deficient of the vitamin its role in bone metabolism is revealed by decrease in density of the long bones and by

slow and irregular healing of fractures. Excessive doses of the vitamin do not, however, accelerate bone healing in persons who have been consuming adequate amounts. Ascorbic acid is required not only for the development of the protein matrix of bone but also for the deposition of bone salt.

2. *Teeth.*—In the teeth the changes are essentially similar to those described as affecting bone formation. The functional activity of odontoblasts is impaired, dentine becomes softened and, together with the tissues of the pulp, is gradually absorbed. No new dentine is formed, but spongy, porous tissue instead. These changes, originally described by Höjer, are so regularly observed in guinea-pigs that they have formed the basis of a method of assay. The teeth of guinea-pigs, however, grow continuously, and it does not follow that similar dental changes occur in man as the result of deficiency. It has been suggested that disease of the gums (gingivitis) in man is due to deficiency of ascorbic acid. It is true that gingivitis, once established, is made worse by a diet grossly deficient in the vitamin and it may be that lack of ascorbic acid in the diet is one of many factors causing this disorder.

3. *Wound Repair.*—Because ascorbic acid is essential for regular formation of the reticular structure of the intercellular matrix of connective tissue it necessarily plays an important role in the laying-down of new connective tissue which is essential to healing. In the absence of ascorbic acid the processes of repair are disturbed, the new connective tissue is weak so that wounds readily break down and the vascularization of scar tissue by the growth into it of new capillaries is much retarded (Fig. 10, 21).

The disorder which occurs in both infants and adults as the result of grossly impaired intake of ascorbic acid is called *scurvy*, to the layman the most familiar of the deficiency diseases. Although in scurvy there is evidence of a general disturbance of bodily function, the patient feeling tired and weak and short of breath, the most characteristic finding is bleeding into the tissues. The typical hæmorrhages in scurvy involve large areas of the deeper tissues but are often subcutaneous also; small pin-point bleedings (petechiæ) are frequently seen in the skin. The hæmorrhagic areas are invariably more prominent in, and may be confined to, parts which are subject to trauma; they therefore tend to appear first in the lower limbs (Fig. 10, 22A). In patients with teeth the gums are swollen, tender and bleeding (Fig. 10, 22B) so that mastication of just those foodstuffs which could effect a cure is most difficult. A group of six human volunteers who were deprived of ascorbic acid during an investigation carried out in Sheffield during the 1939–45 war all developed scurvy within twenty-six to thirty-five weeks. About the same time the gums became swollen and spongy and bleeding occurred in the interdental papillæ. In children bleeding under the periosteum and into joint cavities may also occur and give rise to permanent deformities. With this exception the abnormalities are all corrected within a few days by the administration of ascorbic acid, but it must be pointed out that a diet which leads to the development of scurvy is as a rule deficient in other materials. The adequate treatment of such a case demands therefore a diet rich in all essential factors.

The cause of the bleeding in scurvy is not yet clear: it has recently been

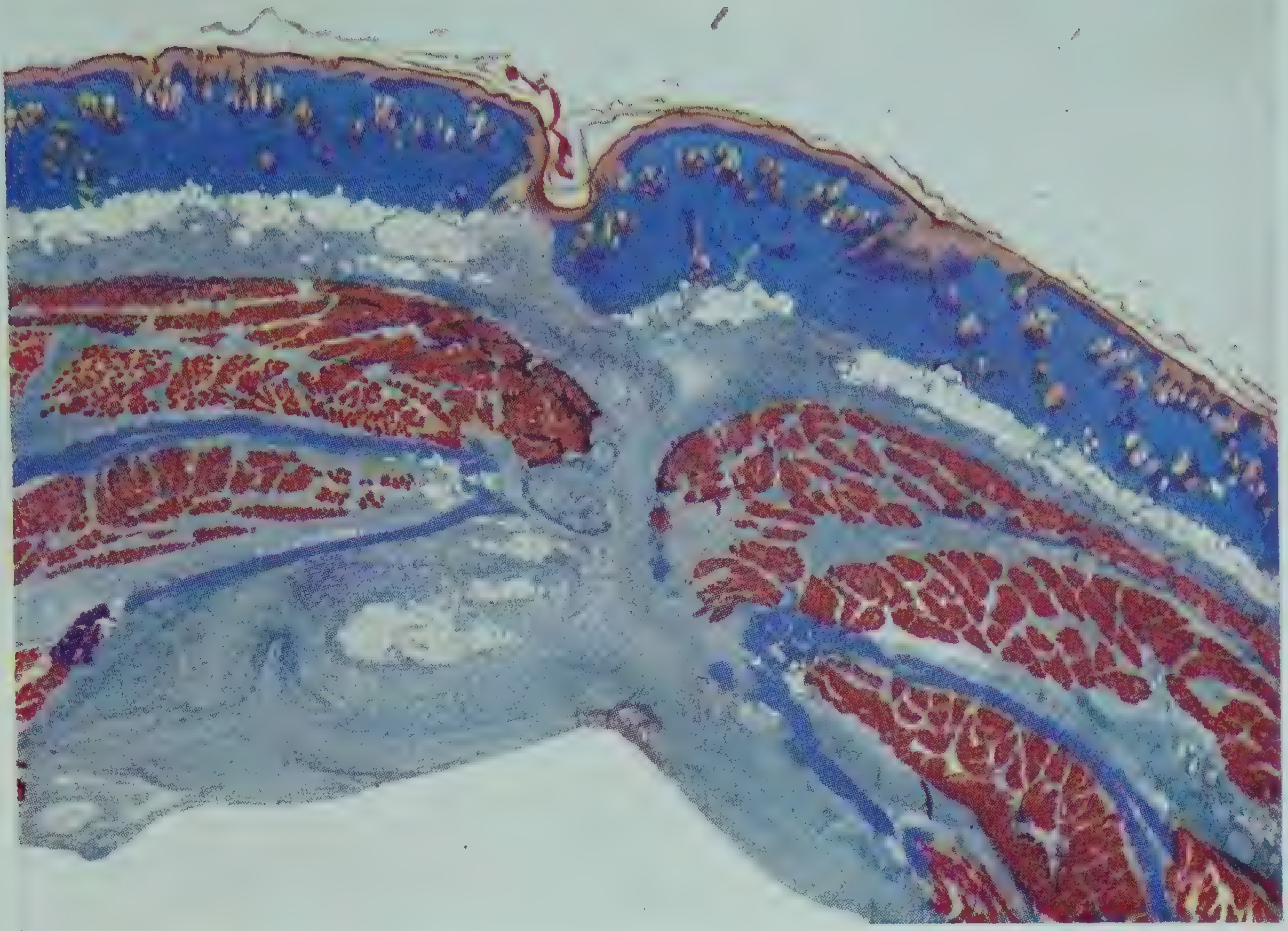


Fig. 10, 21A

Section through a wound experimentally produced six days previously in a normal guinea-pig. This and the section below (× 20) have been stained with the azan method which stains collagen blue; muscle fibres and skin are coloured red. Note the collagen formation throughout the wound, the absence of exudate and the compact nature of the healing process. The tensile strength of this wound was 674 g. per cm.

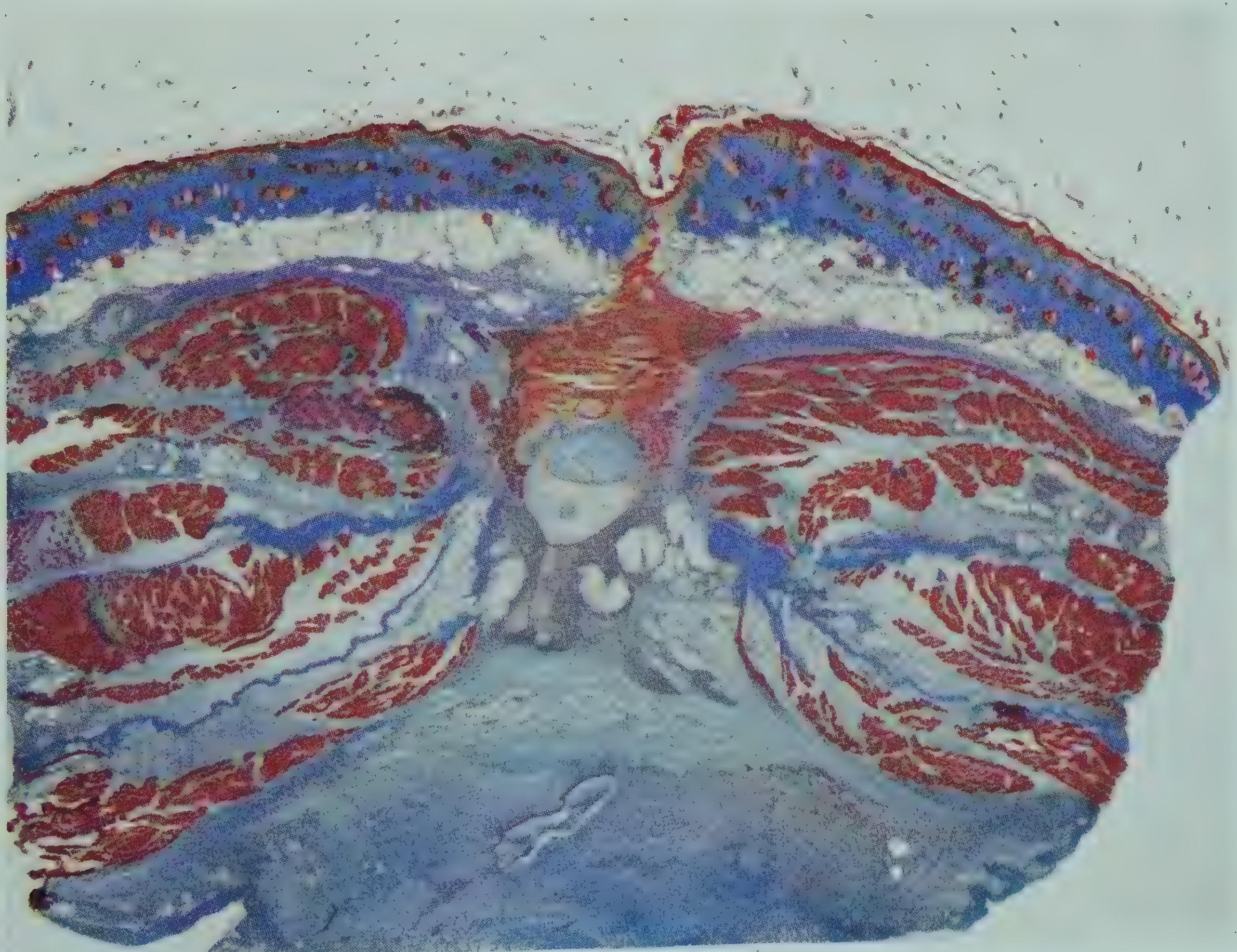


Fig. 10, 21B

A similar wound in a guinea-pig deficient in ascorbic acid. There is still much exudate (red) lying between the muscle fibres; collagen formation is defective and such as is present in the deeper parts of the wound is loose and bulky. Tensile strength, 291 g. per cm.

(J. B. Hartzell & W. E. Stone (1942), *Surg. Gynec. Obst.* **75**, 1.)

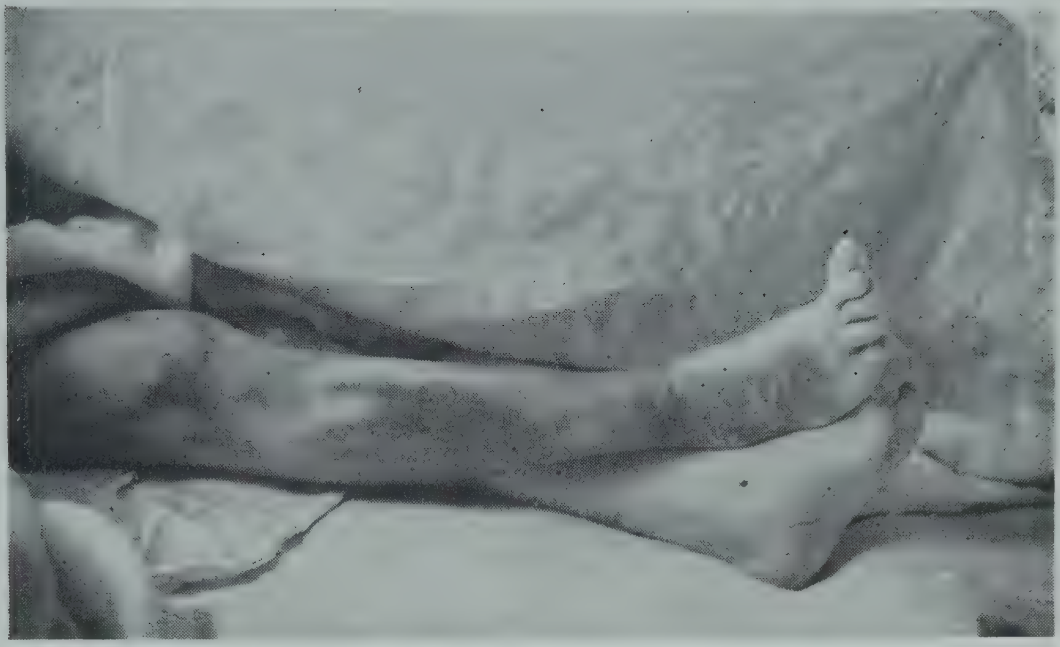


Fig. 10, 22a

Large extravasations of blood into the tissues of the right leg in a man, aged 60, with scurvy.

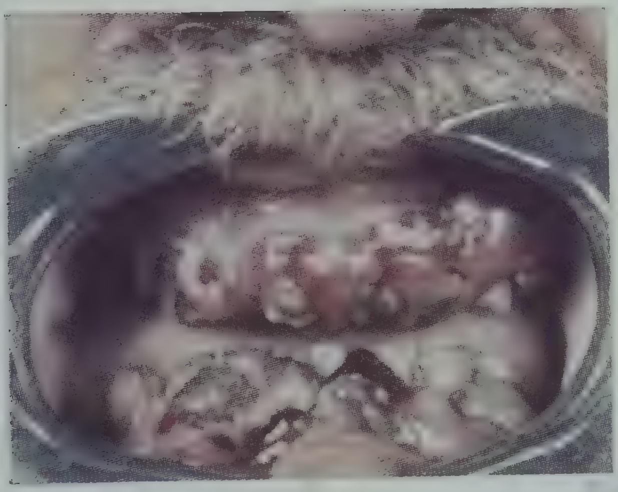


Fig. 10, 22b

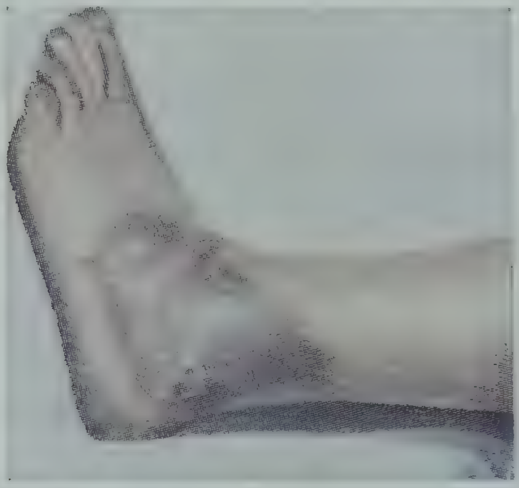


Fig. 10, 22c

Scurvy in a male aged forty-five. Patient, a bachelor, attended clinic because of bleeding of the gums. Fig. 10, 22b shows the maroon spongy gums; the swelling is marked in those of the lower jaw. Fig. 10, 22c illustrates characteristic bruises of the ankle. (H. H. Stones (1951), *Oral and Dental Diseases*, 2nd edition, p. 586. Edinburgh: Livingstone.)

suggested that it comes from the smallest collecting venules. Although frequently the capillaries appear to be more fragile than normal this is by no means always so and tests of capillary strength (resistance or fragility) are of no help in the diagnosis of scurvy. In the scorbutic guinea-pig the arterioles appear to be less sensitive to the injection of adrenaline than those of a normal animal but the relevance of this observation to man is not yet clear.

Ascorbic acid is found in nearly all fruits and vegetables; especially rich sources are rose hips, blackcurrants, parsley, green peppers and the citrus fruits (Table 10, 23). Although potatoes contain relatively smaller amounts they are the most important natural source since they are eaten in quite large amounts by the whole population. It is important to remember when consulting food tables that the amount of vitamin C in a food varies considerably according to climate, soil, maturity, conditions of harvesting, and freshness (Table 10, 24). Losses are partly due to destruction of the vitamin by oxidation and further losses occur during preparation of the food for the table, if the food is finely divided before cooking or if the cooking is unduly prolonged. Much of the vitamin may pass into the cooking water and serious losses occur in food kept warm for long periods before it is served (Table 10, 25). It is thus difficult to estimate the amount of ascorbic acid consumed even if the vitamin content of the raw materials of the diet is known. Canning, deep-freezing and dehydration of foods containing ascorbic acid can be carried out with comparatively small losses.

TABLE 10, 23

Ascorbic Acid in Foods. mg. per 100 g.

Rose hips	500 to 1500	Lemon	20
Paprika	230	Grapefruit	20
Blackcurrants	200	Tomato	20
Parsley	150	Peas—green	12
Kale (cooked)	90	sprouted	20 to 50
Cabbage (raw)	50	Potatoes (uncooked)	5 to 28
Spinach (raw)	50	Lettuce	12
Orange	40		

TABLE 10, 24

The daily intake of ascorbic acid at different seasons of the year in an institution in which potatoes occupied a prominent place in the diet. The intake in mg. per head per day was estimated by chemical analysis of the diet

August	36	January February	28
September	50	March	32
October	48	April May	24
November	55	June	20
December	47		

The new potato crop was lifted during late August and September and stored for the remainder of the year.

As with many other vitamins, any attempt to assess the daily requirement can be no more than provisional. The daily allowance of ascorbic acid recommended by the Food and Nutrition Board of the U.S. National Research Council is 75 mg. per day which is essentially the amount necessary to maintain saturation. However, in the Sheffield experiment already referred to, it was found that the daily dose of ascorbic acid necessary to prevent scurvy was 10 mg. or somewhat less. A daily intake of 10 mg. per day is therefore the minimum protective allowance. To allow for individual variations in requirements and to provide a margin of safety, three times this dose, that is 30 mg. per day, has been suggested as the daily requirement for a healthy adult. Whether such an allowance is optimal cannot yet be stated but there is no evidence that any advantage is to be gained by a higher intake.

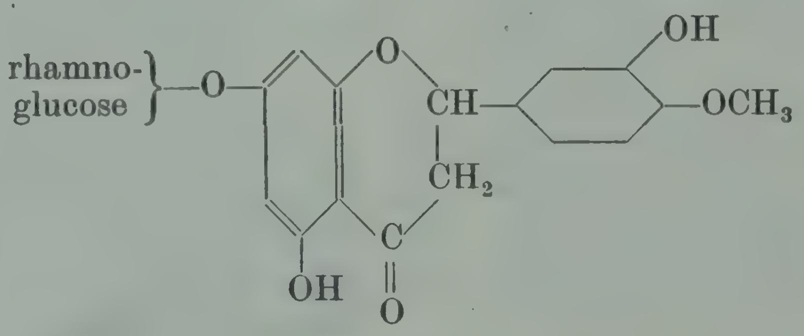
Table 10, 25

Loss of ascorbic acid from cooked cabbage in relation to the time of keeping hot

10 minutes	14 per cent. loss	40 minutes	50 per cent. loss
30 minutes	25 per cent. loss	60 minutes	75 per cent. loss

Vitamin P.—Vitamin P is the name given by Szent-Györgyi in 1936 to a factor present in red peppers and lemon peel which was believed to increase the strength of the cutaneous capillaries in man. It has now been established that a dietary factor (vitamin P) is required for the maintenance of capillary strength (resistance) in the guinea-pig but deficiency symptoms have never been reported.

The constitution of vitamin P is not known and its very existence is questioned by some. It is probable that the term comprises a large group of substances of different degrees of activity many of which are chemically related to the flavanone glycoside, hesperidin. Such compounds are found widely distributed in plants; blackcurrants, grapes, rose hips and citrus fruits are rich natural sources.



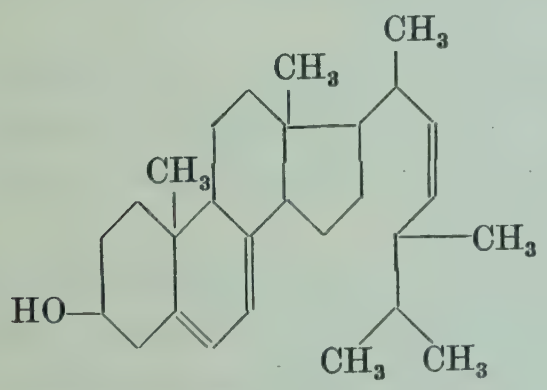
hesperidin

VITAMIN D

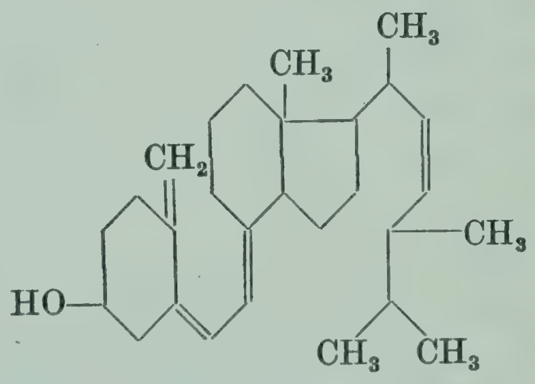
The first recorded account of the successful use of cod-liver oil in the treatment of *rickets* was published in 1782 but, although cod-liver oil was used on the Continent for many years as an antirachitic agent, this important clinical observation was largely overlooked. It was not until Mellanby produced experimental rickets in puppies in 1918 that it began to be accepted that rickets is a deficiency disease. This opinion was for some time opposed by Findlay and others in Glasgow, who thought that lack of fresh air, exercise and sunlight was the cause of rickets. The two apparently opposing views were reconciled by Harriet Chick who went to Vienna in 1919 to investigate rickets, which had become very prevalent in Central Europe as a result of the food shortages following the 1914–18 War. She found that either cod-liver oil or sunlight (or artificial sunlight produced by a mercury-vapour lamp) cured rickets. It is now known that animals and man can manufacture the antirachitic vitamin in their own bodies under the influence of ultraviolet light from the sun, or obtain it at second-hand by eating the products of other animals (dairy produce) that have already synthesized it.

The name vitamin D embraces a number of chemically related compounds all of which possess the property of preventing or curing rickets in greater or less degree. The two most important are vitamin D_2 (ergocalciferol) and vitamin D_3 (cholecalciferol). Ergocalciferol, which does not occur naturally, is the active compound with which most of the experimental work has been done. It is produced artificially by the action of ultraviolet light on ergosterol (p. 25), which is thus a provitamin. All the provitamins of the D group are sterols which, like ergosterol, acquire antirachitic properties by a slight rearrangement of molecular structure when exposed to ultraviolet light. However, many other compounds, some of them toxic, are formed in the process of converting the provitamin to the physiologically active substance.

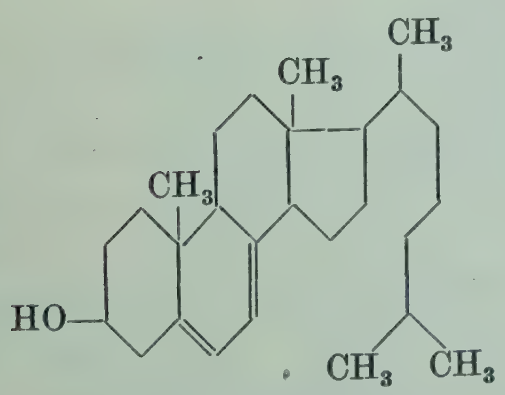
The provitamins D are all steroid derivatives (p. 23), the structural formulæ for vitamin D_2 and D_3 being as follows:



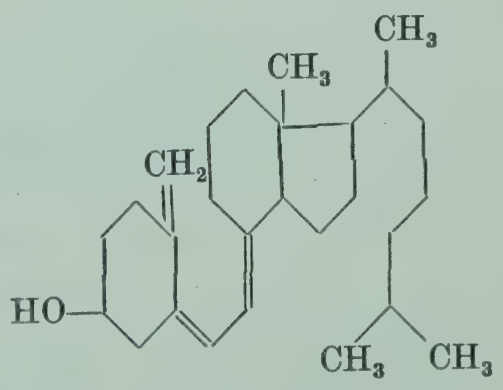
ergosterol



vitamin D_2 (ergocalciferol)



7-dehydrocholesterol



vitamin D_3 (cholecalciferol)

Vitamin D_3, the naturally occurring antirachitic compound of mammals, was first isolated in 1936 from fish-liver oil. It is produced by the irradiation by the sun of 7-dehydrocholesterol present in the secretions on the surface of the skin. The reaction must occur either actually on the surface of the skin or immediately below it, since ultraviolet light can penetrate only a short distance. In animals the irradiated material is often licked off the fur: in man it is presumably absorbed into the blood.

Vitamin D is not readily destroyed by heat; its thermostability enabled McCollum to distinguish between it and vitamin A. The vitamin D potency of materials is measured with reference to the International Standard which is a solution of irradiated ergosterol. The international unit (i.u.) is the activity of 1 mg. of the International Standard, which has been found to be equivalent to that of 0·025 μg. of pure crystalline vitamin D_3. Vitamins D_2 and D_3 are equal in antirachitic potency in the rat and in man but in chicks vitamin D_3 is considerably the more active antirachitic compound.

In spite of an enormous amount of research the role of vitamin D in metabolism has not been fully elucidated and therefore the biochemical lesion produced by lack of the vitamin has not been defined. At the present time the actions of vitamin D are best described by considering what happens when the vitamin is administered to a deficient animal (not a rat). Both direct and secondary effects can be recognized. The most important direct effect is an improved absorption of calcium from the intestine. It is not known how this is produced. It may be a direct influence upon the cells of the mucous membrane or the result of the reduction in the pH of the intestinal contents that follows vitamin D administration. The vitamin also acts directly on the bones since an early effect of its administration to a deficient animal is a rapid increase in the citric acid content of bone and, later, an increase in ash content. In vitamin D deficiency there is a decrease in the percentage of inorganic material in the bone and as the vitamin deficiency is corrected the percentage of inorganic salt increases.

Among the secondary effects produced is an improved absorption of phosphate from the intestine, an increase in the phosphoric esters in the red blood cells and an increased absorption of phosphate by the renal tubules.

The early signs of vitamin D lack in man are an increased elimination of phosphate in the urine and a reduction in the serum level of inorganic phosphate. An early and constant elevation of the concentration of the enzyme alkaline phosphatase in the blood reflects increased osteoblastic activity in the bones.

Rickets is a disorder of which the main feature is a failure of calcium salts to be deposited promptly in the newly formed bone matrix. The delayed ossification is not due to a defect in the bones themselves because bone and cartilage from rachitic animals contain phosphatase and are able to lay down bone salts if they are immersed in normal serum or in suitable chemical media, and the injection of calcium salts and phosphates into rachitic animals is followed by deposition of bone. The most striking changes in rickets are to be found in the skeleton. During normal bone growth the cells in the cartilage at the growing ends of the diaphysis (shaft) degenerate and are replaced by osteoblasts, the cells responsible for the formation of the bony matrix. The osteoblasts are finally transformed into bone cells as the tissue is invaded by capillaries in which are carried the bone-forming salts. In the epiphysis the cartilage cells proliferate so that growth of cartilage keeps pace with its conversion

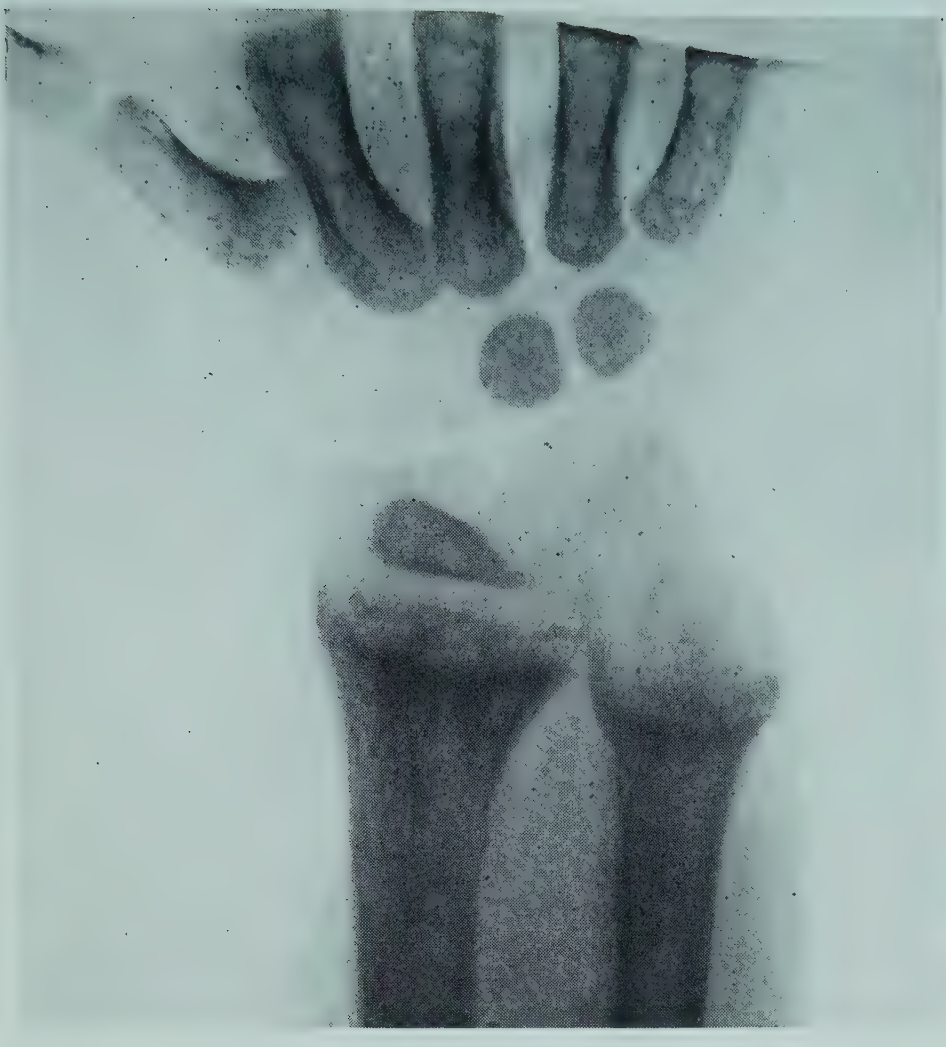


FIG. 10, 26A

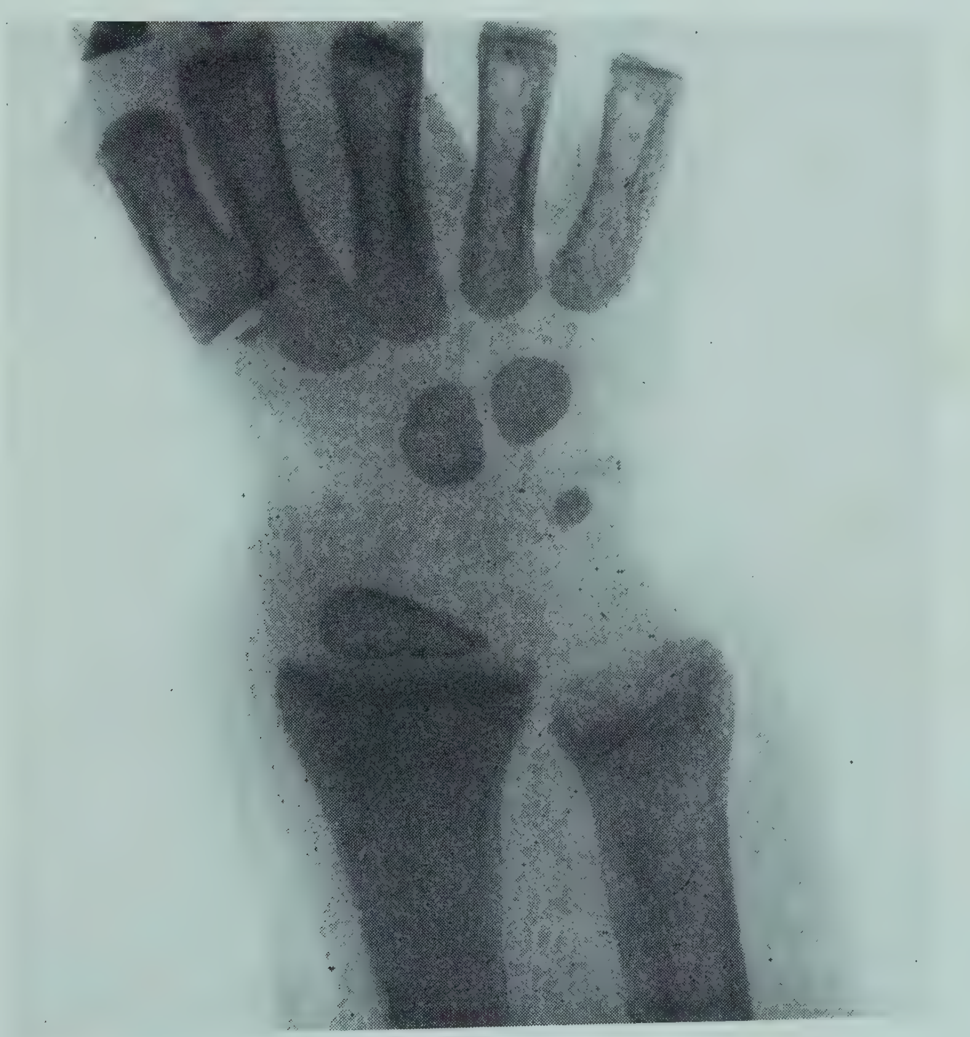


FIG. 10, 26B

Radiological appearances before and after the administration of 600,000 i.u. of calciferol to a child with active rickets. In the upper photograph note the 'ragged' appearance of the end of the diaphysis produced by the irregular calcification and degenerating cartilage. The shafts of the bones show rarefaction and the outlines are irregular as the result of partially calcified osteoid tissue. 600,000 i.u. calciferol were given after the first photograph was taken. Second film taken three weeks later.

(By courtesy of Stanley Graham.)

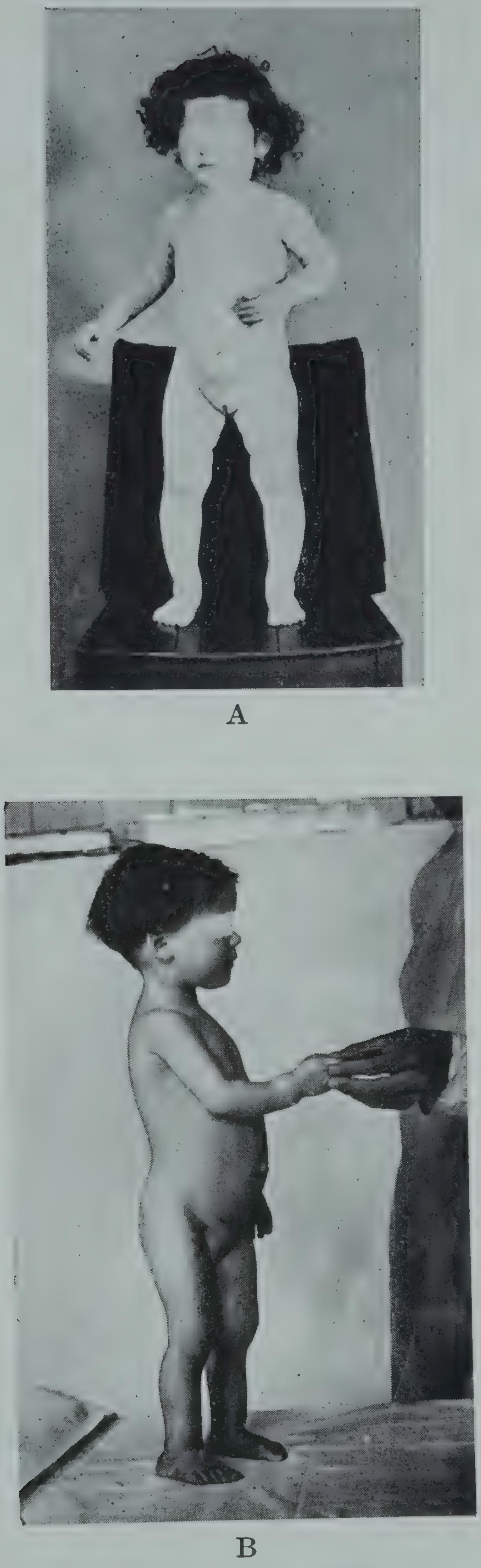


Fig. 10, 27

Active rickets. Note the bowing of the legs and also the swelling near the wrist, the result of excessive osteoid tissue at the growing ends of the bones of the forearm. The child in B, with more severe rickets, shows deformities of the lower end of the tibiæ and bowing of the radius and ulna.

(By courtesy of Stanley Graham.)

to bone. In vitamin D deficiency the formation of cartilage continues but osteoblasts are no longer produced in an orderly way and bone formation becomes irregular and finally ceases. The band of cartilage at the end of growing long bones becomes much wider than normal and, since cartilage is transparent to X-rays while bone is not, the characteristic radiological picture of rickets is produced (Fig. 10, 26A). The expanded cartilage produces swellings at the costochondral junctions ('rickety rosary') and similar swellings at the ends of the long bones, particularly the lower end of the radius (Fig. 10, 27). In more prolonged deficiencies there is considerable reabsorption of bone material so that the bones already laid down become weaker. The process of bone modelling which normally maintains the shape of the bones is disturbed and the bones bend under the child's weight. These changes at the epiphyses are very rapidly reversed when vitamin D is given and the bones may become ossified in the abnormal shape acquired during the active stage of the disease. The deformities are then permanent (Fig. 10, 27). From what has been said it will be clear that rickets is a disease which particularly affects growing animals. In adults deficiency of vitamin D leads to *osteomalacia*, in which there is reabsorption of bone already laid down; this condition may occur in pregnant women if their absorption of minerals is insufficient to meet the needs of the fœtus.

It has often been held that inadequate intake of vitamin D is a factor in producing dental decay. Experiments have shown that rachitic puppies have hypoplastic teeth with poorly developed enamel and dentine. The addition of supplements of vitamin D to the diets of children diminishes the spread of dental decay already present and reduces the amount of new caries forming subsequently. However, it is evident that vitamin D can neither prevent nor arrest dental caries so it cannot be the only factor involved in this disease. It should be emphasized again that both vitamin A and ascorbic acid have important roles in the development and maintenance of healthy teeth.

Because the quantitative estimation of vitamin D involves complex and lengthy biological assay there is little precise information about the storage of the vitamin in tissues. In the rat and in the dog bile is essential for the absorption of vitamin D from the intestine. Absorption is hindered by administration of liquid paraffin since the vitamin, like other fat-soluble vitamins, is soluble in the unabsorbable mineral oil and is therefore lost to the body.

Animal experiments have shown that the effects of a diet lacking vitamin D are considerably modified by the constitution of the experimental diet. Rats do not develop rickets unless the relative proportions of calcium and phosphorus are grossly abnormal but it is likely that children living on vitamin-deficient diets would become rachitic even when the Ca: P ratio of the diet was normal. The proportion of cereal in the diet, especially of wheat-germ and oatmeal, has an important bearing on the development of rickets, since cereals contain phytic acid (inositol hexaphosphate) the phosphorus of which is not available for absorption (p. 178). In addition phytic acid combines with calcium in the intestine and renders this unavailable. The custom of taking milk with cereal or porridge receives physiological justification since it provides an extra source of calcium.

The only important sources of vitamin D in the human dietary are milk, cream, eggs, butter and 'vitaminized' margarine (80 to 100 i.u. per oz. by law), and fatty fish (Table 10, 28). Vegetables, fruit and cereals contain negligible amounts.

TABLE 10, 28

Vitamin D in Foods. i.u. per 100 g.

Food	i.u.	Food	i.u.
Tunny-liver oil	85,000 to 25,000,000	Ox liver	40 to 50
Halibut-liver oil	20,000 to 400,000	Egg yolk	150 to 500
Cod-liver oil	2000 to 30,000	Butter	8 to 100
Herring-body oil	10,000 to 20,000	Cow's milk (raw)	0·3 to 3·8
Kipper oil	10,000 to 15,000	Margarine (vitaminized)	260 to 330
Fresh herring	800		
Canned herring	160		

The vitamin D content of all farm produce depends, however, on the diet and on the exposure to sunlight of the cows and hens. The vitamin is added as irradiated ergosterol to margarine during manufacture. Milk may be fortified or enriched by irradiation of the milk, or of the cow, or by adding the vitamin directly to the milk. There is good experimental evidence that the vitamin when taken in milk is a more efficient antirachitic than when it is consumed in any other form, for example as a concentrate. The fish-liver oils are valuable sources of vitamin D and, as we have seen, are rich also in vitamin A. Cod-liver oil is still one of the best means of administering vitamin D and it has the advantage that it contains vitamin D_3, the natural form of the vitamin. Vitamin D is a relatively stable substance so that losses occurring during the storage, preparation, canning, or dehydration of foods are probably small.

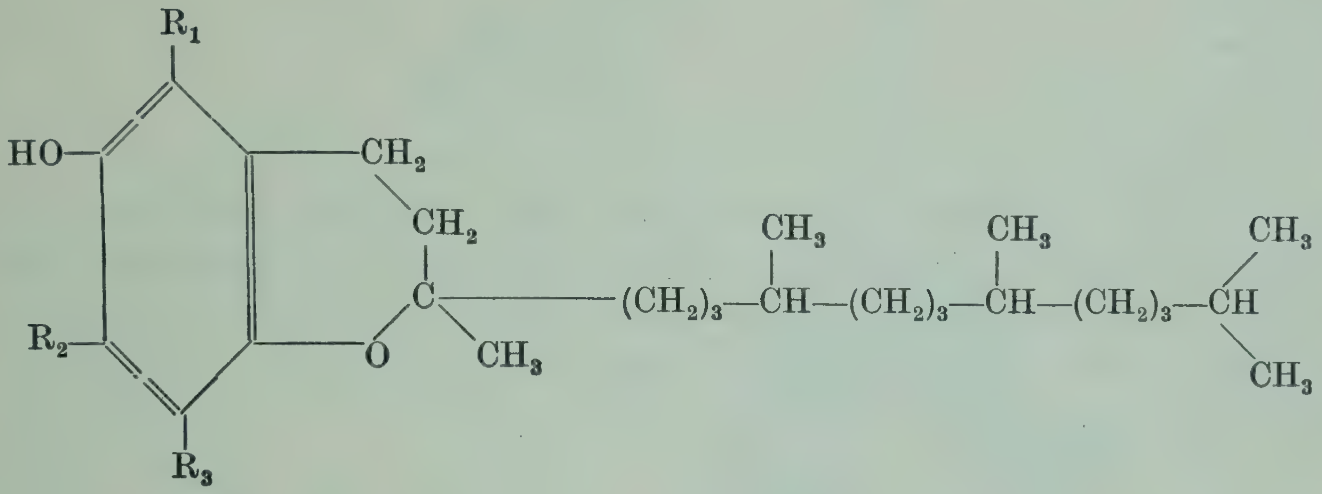
Because of the low vitamin D potency of most foods it is important that man should benefit by the natural method of obtaining the vitamin—as the result of the sun's ultra-violet radiation upon the sterols in the skin. It is probable that the enormous reduction in the incidence of rickets during the last thirty years is due as much to changed ways of bringing up children as to improvements in their diets.

TOCOPHEROL (VITAMIN E)

Vitamin E is the name given to a number of chemical compounds possessing in greater or less degree the same biological function. The most active of these compounds is α-tocopherol, the dextro-rotatory form of which is somewhat more potent than its optical isomer. The corresponding β- and γ-tocopherols, with two CH_3 groups attached to the aromatic ring instead of three as in α-tocopherol, are less active. Although the effects of vitamin E deficiency had been known in rats as early as 1922 the vitamin was only finally isolated from wheat-germ oil by Evans, Emerson and Emerson in 1936, and has since been synthesized.

The usual chemical methods of estimating vitamin E are based either upon measurements of the absorption at 294 mμ or upon its reducing properties; the reduction of gold chloride or of ferric chloride are the most commonly employed reactions.

The biochemical lesion of vitamin E deficiency has not been identified. The vitamin may be concerned with the oxygen uptake of tissues because vitamin E deficiency in young rats is characterized by an increased oxygen consumption, probably confined to the skeletal muscles, and this can rapidly be reduced by injection of the vitamin. *In vitro*, vitamin E behaves as an *antioxidant* (p. 21) protecting both vitamin A and carotene from oxidation. It tends also to delay the rancidity of fats. In these processes vitamin E is itself oxidized and inactivated.



α-tocopherol, R_1, R_2 and $R_3 = CH_3$; β-tocopherol, R_1, $R_3 = CH_3$, $R_2 = H$; γ-tocopherol R_2, $R_3 = CH_3$, $R_1 = H$.

The most characteristic changes which follow deprivation of vitamin E in rats occur in the reproductive organs, but prolonged deficiency causes changes in many other tissues. Female rats do not, in the early stages of deprivation of the vitamin, lose the power of conception when mated with a normal male and implantation of the ovum occurs normally. After about the first week of gestation, however, fœtal growth becomes retarded, apparently because of maldevelopment of the fœtal mesoderm, so that the fœtuses die at the end of the second week and are resorbed (Fig. 10, 30). If vitamin E is given subsequent pregnancies proceed normally. Prolonged deprivation of vitamin E in the adult female rat leads to disturbance of the œstrous cycle and loss of the power of conception. In male rats on diets lacking vitamin E there is progressive atrophy of the testes with degeneration of the sperm-forming cells leading to permanent sterility which cannot be reversed by restoring the vitamin to the diet. Widespread degenerative changes in skeletal muscle leading to actual paralysis (nutritional muscular dystrophy) have been shown to occur in rats, guinea-pigs and rabbits.

There is little evidence that a deficiency syndrome exists in man and the relation of sterility to vitamin E deficiency in the human species is doubtful, although the treatment of habitual abortion with tocopherol has met with some success. The absorption of tocopherol from the alimentary tract is defective when the absorption of fat is impaired (p. 274) and the plasma concentration, normally about 1 mg. per 100 ml., tends to fall. Considerable stores of tocopherol are probably present in man and certainly in the rat.

Vitamin E from natural sources is remarkably stable. It occurs in small amounts in most vegetables, in dairy produce and in meat. The richest sources are the vegetable oils, of which wheat-germ oil is the most potent (Table 10, 29), but fish-liver oils are almost devoid of it.

TABLE 10, 29

Tocopherol (Vitamin E) in Foods. mg. per 100 g.

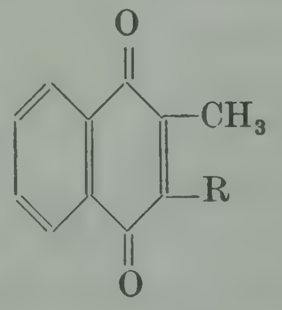
Wheat-germ oil	520	Meat	0·5 to 1·0
Wheat germ	27	Liver	2 to 3
Bread—brown	2	Vegetables	0·1 to 5·0
white	1·5	Lettuce (dry)	55

Because of the difficulty of estimating the tocopherols in food little is known about requirements. It may be pointed out, however, that the replacement of wholemeal by flour from which the wheat-germ has been removed must have resulted in a lower intake of the vitamin for the majority of the population.

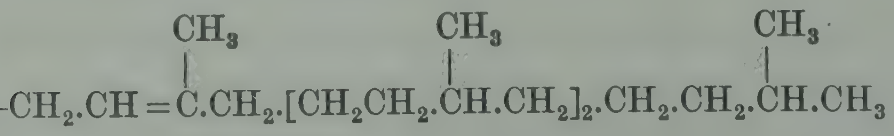
VITAMIN K

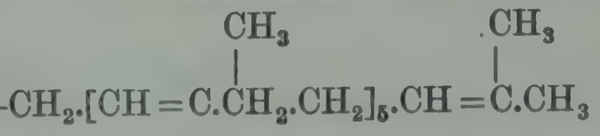
Between 1929 and 1933 work on a bleeding disease which develops in chickens on diets lacking in greenstuffs and poor in fats led Dam to the discovery of the fat-soluble vitamin K. Deficiency of this substance results in impaired coagulability of the blood. In man it is doubtful whether a diet deficient in vitamin K can ever lead to impaired blood coagulability.

There are at least two naturally occurring forms of the vitamin. Vitamin K_1 is synthesized in the green parts of plants, vitamin K_2 by certain micro-organisms. A large number of active substances has now been synthesized, several of which are more potent biologically than either of the naturally occurring vitamins. All these active compounds contain the 2-methyl-1 : 4-naphthoquinone structure.



In 2-methyl-1 : 4-naphthoquinone R is H

In vitamin K_1 : R is 
$-CH_2.CH=\overset{CH_3}{\overset{|}{C}}.CH_2.[CH_2CH_2.\overset{CH_3}{\overset{|}{C}}H.CH_2]_2.CH_2.CH_2.\overset{CH_3}{\overset{|}{C}}H.CH_3$

In vitamin K_2 : R is 
$-CH_2.[CH=\overset{CH_3}{\overset{|}{C}}.CH_2.CH_2]_5.CH=\overset{CH_3}{\overset{|}{C}}.CH_3$

Both vitamins have been isolated, K_1 from alfalfa, K_2 from putrefying fish meal, and vitamin K_1 has been synthesized. The bacterial flora normally

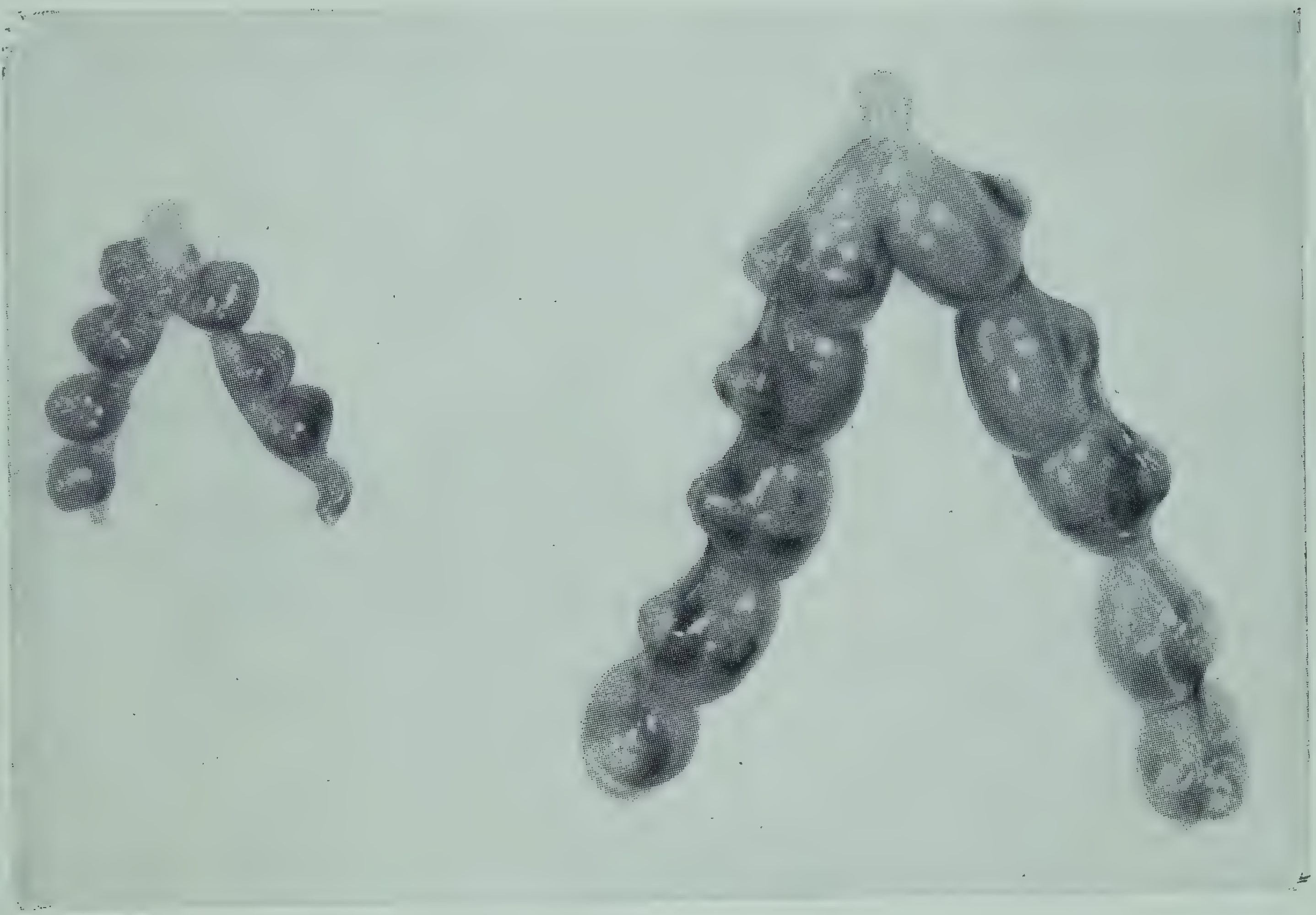


FIG. 10, 30

The pregnant uterus from a vitamin-E-deficient rat (left) is compared with that from a rat on a normal diet (right). The embryos in the deficient rat are fewer in number and their development has been arrested. (From F. Bicknell & F. Prescott, *Vitamins in Medicine*, p. 711, Fig. 177. London: Heinemann.)

(By courtesy of Margaret Wright.)

present in the human intestine can synthesize vitamin K and probably for this reason deficiency symptoms have never been produced in man or other mammals solely as the result of diets lacking in the vitamin.

In vitamin K deficiency the process of blood coagulation (p. 364) is disturbed as the result of a reduction in prothrombin and factor VII in the blood. A useful clinical index of this disturbance is the 'prothrombin time' (p. 366) which, normally 12 to 15 seconds, may increase to over 30 seconds. This impaired coagulability of the blood may in certain circumstances, for example after a surgical operation, actually lead to hæmorrhage (Fig. 10, 31). Synthesis of prothrombin takes place mainly in the liver so that, in diseases in which the liver is extensively damaged, dangerously low levels of plasma prothrombin and factor VII may be found even when adequate amounts of vitamin K are taken in the diet. In these conditions the administration even of large amounts of the vitamin may lead to little improvement in coagulation.

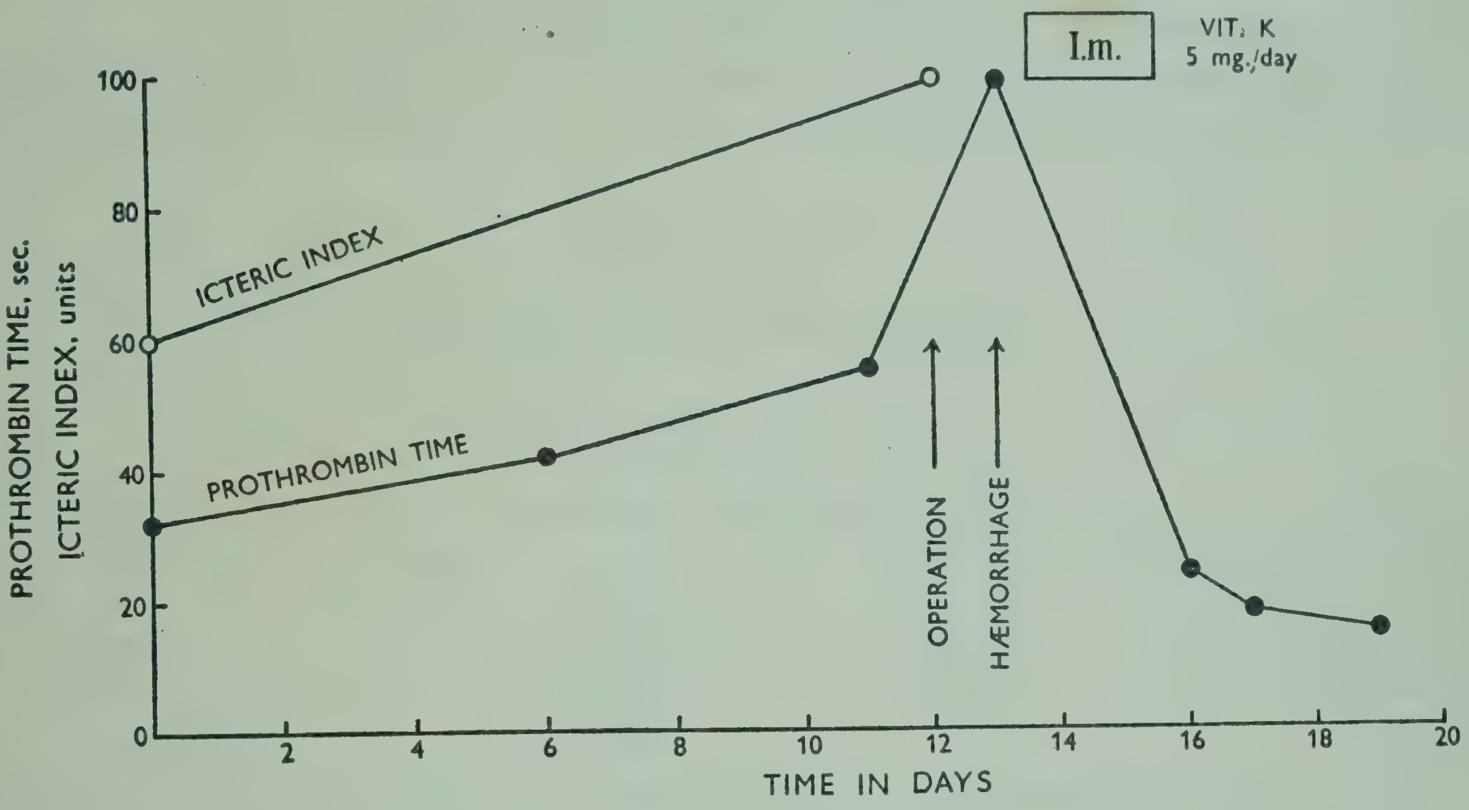


FIG. 10, 31

Showing a progressive increase in prothrombin time (p. 366) and a rise in the icteric index in a patient with obstructive jaundice. After operation the prothrombin time rose sharply and a small hæmorrhage occurred. This was arrested and the prothrombin time restored to normal levels by injection of a synthetic vitamin K analogue. The icteric index is a measure of the concentration of bile pigment in blood serum.

Vitamin K is fat-soluble and its absorption from the alimentary tract therefore depends upon bile salts. In obstructive jaundice (p. 415), in which the amount of bile salt reaching the intestine may be much reduced, vitamin K, whether taken in the food or synthesized by the intestinal flora, is not absorbed in adequate amounts so that the prothrombin time is prolonged and bleeding may occur. Likewise, when the absorption of fat from the intestine is defective the absorption of the fat-soluble vitamins is also impaired and a tendency to hæmorrhage may develop. Bleeding due to a combined deficiency of prothrombin and factor VII is most frequently seen in new-born infants. This may be caused by reduced permeability of the placenta to vitamin K, inadequate maternal intake of the vitamin, or perhaps because the intestine does not for some time develop the bacterial flora which synthesizes vitamin K.

Except when due to liver disease, the disturbance of coagulation can be corrected by synthetic analogues of the vitamin, preferably administered by injection. If the vitamin is given by mouth it is necessary for the patient to take bile salts also, especially in those disorders in which the vitamin deficiency has been brought about by a failure of fat absorption due to the absence of bile salts from the intestinal content. The development of the hæmorrhagic disorder in the new-born can largely be prevented by giving vitamin K or a synthetic analogue to the infant and a normal prothrombin time in the infant's blood ensured by the administration of these substances to the mother shortly before delivery.

The best natural sources of vitamin K are the green parts of plants, especially those of alfalfa, spinach, kale and cabbage (Table 10, 32).

TABLE 10, 32

Vitamin K in Various Materials

Dam units * per 100 g.

Alfalfa	20,000 to 40,000	Tomato—green	10,000
Putrefied fish meal	90,000	ripe	5000
Cabbage leaves	55,000	Liver	5000
Spinach leaves	55,000	Carrots	1000
Cauliflower	40,000	Potatoes	1000
Fæces	30,000	Milk	negligible

* The potency of the material to be tested is referred to that of a standard preparation of dried spinach one gram of which is arbitrarily assigned a value of 500 Dam units.

Vitamin K_2 is found in many micro-organisms and most putrefying animal or vegetable matter therefore contains large amounts.

The dietary requirements of man are unknown.

REFERENCES

BICKNELL, F. & PRESCOTT, F. (1953). *The Vitamins in Medicine*, 3rd ed. London: Heinemann.

EDDY, W. H. & DALLDORF, G. (1944). *The Avitaminoses.* 3rd ed. London: Baillière, Tyndall & Cox.

HARRIS, L. J. (1951). *Vitamins—A Digest of Current Knowledge.* London: Churchill.

HARRIS, L. J. (1955). *Vitamins in Theory and Practice*, 4th ed. Cambridge University Press.

LOWRY, O. H. (1952). Biochemical evidence of nutritional status. *Physiol. Rev.* **32,** 431.

MEIKLEJOHN, A. P. (1953). The physiology and biochemistry of ascorbic acid. *Vitamins and Hormones,* **11,** 61–89.

NOVELLI, G. D. (1953). Metabolic functions of pantothenic acid. *Physiol. Rev.* **33,** 525–543.

Recent research on vitamins (1956). *Brit. med. Bull.* **12,** 1–90.

REED, L. J. (1953). Metabolic functions of thiamine and lipoic acid. *Physiol. Rev.* **33,** 544–559.

ROBINSON, F. A. (1951). *The Vitamin-B Complex.* London: Chapman & Hall.

SEBRELL, W. H. & HARRIS, R. S. (1954). *The Vitamins,* 3 vols. New York: Academic Press.

SMITH, D. A. & WOODRUFF, M. F. A. (1951). Deficiency diseases in Japanese prison camps. *Spec. Rep. Ser. med. Res. Coun., Lond., No.* 274. London: H.M.S.O.

Snell, E. E. (1953). Summary of known metabolic functions of nicotinic acid, riboflavin and vitamin B_6. *Physiol. Rev.* **33,** 509–524.

Symposium on the role of vitamins in metabolic processes (1953). *Proc. Nutrit. Soc.* **12,** 83–128.

Symposium on vitamin A (1951). *Brit. J. Nut.* **5,** 94.

The biochemistry of vitamin B_{12}. *Biochem Soc. Symp., No.* 13. Cambridge University Press.

Vitamin A requirements of human adults (1949). *Spec. Rep. Ser. med. Res. Coun., Lond., No.* 264. London: H.M.S.O.

Vitamin C requirements of human adults (1953). *Spec. Rep. Ser. med. Res. Coun., Lond., No.* 280. London: H.M.S.O.

Vitamins and Hormones. New York: Annual publication of Academic Press Inc.

CHAPTER 11

ENERGY EXCHANGE

THE main function of organic foodstuffs in the animal body is to supply energy. Energy is required for the maintenance of body temperature, for the muscular movements of the heart and of respiration, for the performance of external work, and for the various synthetic processes which occur in the tissues, for example, the building up of proteins and hormones.

This energy is derived ultimately from the oxidation of foodstuffs but the amount obtained varies considerably according to the actual material which is being utilized. The potential energy of a foodstuff can be determined by measuring the amount of heat produced when the foodstuff is burned; for this purpose a *bomb calorimeter* is employed (Fig. 11, 1). It consists essentially of a stout metal case which can resist internal pressure. A weighed amount of the foodstuff is placed in the bomb which is filled with oxygen under pressure. The bomb is then placed in a can of water, surrounded in turn by a vessel insulated with cork to reduce heat loss. The foodstuff is ignited electrically and the amount of heat produced is measured by determining the rise in temperature of the water surrounding the bomb.

The unit of heat used in work of this sort is the kilocalorie (kcal). One kilocalorie (calorie or large calorie) is the amount of heat required to raise the temperature of one kilogram of water from 15° to 16° C and is equal to 1000 gram-calories or small calories. The energy derived in the bomb calorimeter from the complete combustion of such a mixture of carbohydrates as occurs in the diet has been found to be 4·1 kilocalories per gram. The figures for fat and protein of average composition are respectively 9·3 and 5·4 kilocalories per gram.

As the result of combustion in the bomb calorimeter protein is oxidized to carbon dioxide, water, oxides of nitrogen and of other elements, but in the body protein is incompletely oxidized, part of it being excreted as urea which still has an appreciable potential energy value. When the potential energy of the urea excreted is subtracted from the figure for the complete oxidation of protein in the bomb calorimeter a physiological value of 4·1 kilocalories per gram is obtained for protein.

Rubner's heat of combustion values or the physiological energy values of the foodstuffs are as follows:

1 g. carbohydrate yields	4·1 kcal
1 g. fat yields	9·3 kcal
1 g. protein yields	4·1 kcal

However, in practical dietetics it is convenient to use rounded-off Atwater factors 4, 9, 4 for carbohydrate, fat and protein respectively. These lower figures take account of digestive losses; when Rubner's values are reduced to allow for these losses they become approximately the same as the Atwater figures.

Since it is easier to measure the volume of a gas than to measure heat production, an apparatus called the *oxycalorimeter* (Fig. 11, 2) has been devised. Although it is not so accurate as the bomb calorimeter the estimations can be carried out more quickly. The principle is similar to that employed in the Roth-Benedict apparatus described on p. 158. A measured amount of food is held in a combustion chamber connected to a gasholder or spirometer containing oxygen. The system is completely shut off from the air and arrangements are made to absorb the carbon dioxide produced; the volume of oxygen used in the combustion is measured from the fall of the spirometer. Before this can be converted into the actual energy evolved it is necessary to know how many kilocalories are produced when a litre of oxygen is consumed. Since the heat liberated per litre of oxygen used in the combustion of foodstuffs (see Table 11, 3) varies from 4·5 to 5·0 kilocalories the use of a mean value of say 4·8 will result in a *maximum* deviation

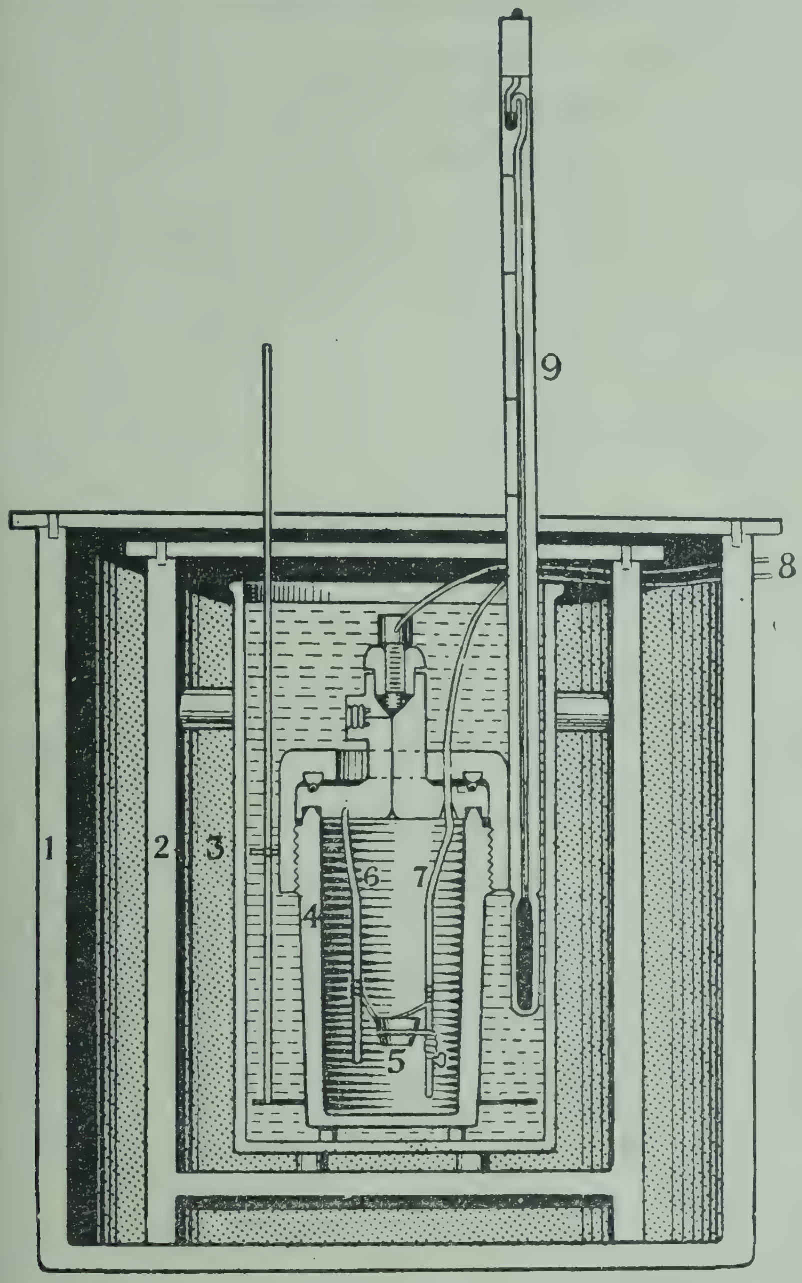


FIG. 11, 1

The Bomb Calorimeter

The bomb calorimeter consists of three vessels —(1), (2) and (3)—one inside the other. The innermost vessel (3) contains water in which is immersed the bomb (4) containing the crucible (5) with the foodstuff to be burned. The foodstuff is ignited by means of the leads (6) and (7) which pass out of the casing at (8). Combustion of the material in the bomb is followed by a rise in temperature of the water surrounding it, as measured by the thermometer (9).

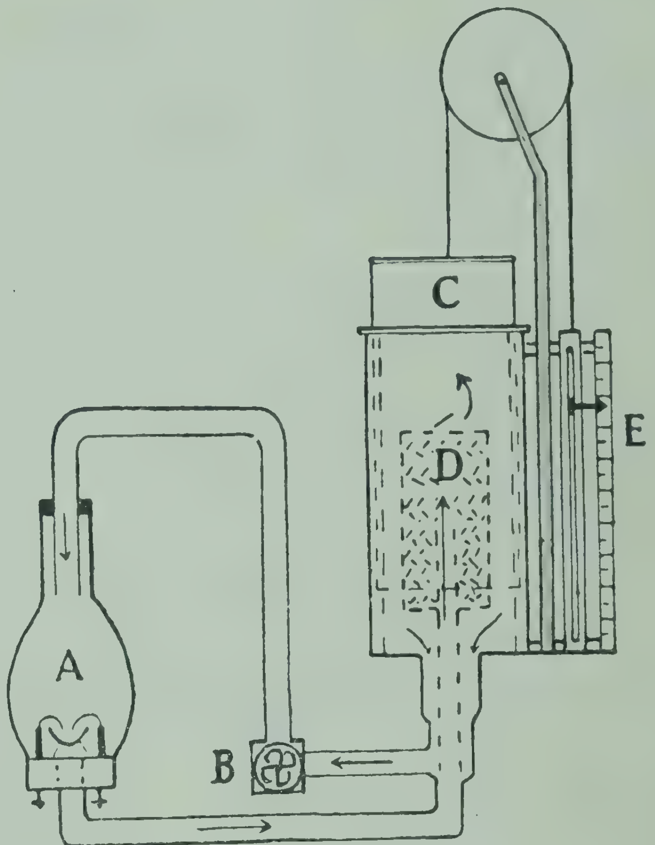


FIG. 11, 2

The Oxycalorimeter

Diagram of oxycalorimeter for determining energy values of foods, etc. (A) Combustion chamber for nickel crucible and electrical connexions. (B) Rotary blower for rapid circulation of the gases. (C) Spirometer bell with oxygen in water seal. (D) Soda-lime container. (E) Scale in millimetres for measuring the oxygen used. (*After* Cruickshank.)

of only 5 per cent. from the true value obtained in the bomb calorimeter. If the nature of the substance burned is known then a better approximation can be used, for example if it is fatty the factor 4·7 would be used. This apparatus is used for the determination of the approximate value of actual diets—Benedict and Fox give as an example a salmon sandwich.

By a similar principle the energy output of an animal may be found by measuring its oxygen consumption during a measured time and calculating from the result the amount of heat produced. Here again it is necessary to know the relationship between oxygen consumption and heat production. This depends on the type of food being oxidized, and in order to understand this point we must briefly consider the *respiratory quotient* (R.Q.) which is by definition the *volume* of carbon dioxide produced divided by the *volume* of oxygen used in the same time. When carbohydrate is oxidized according to the equation

$$C_6H_{12}O_6 + 6O_2 = 6CO_2 + 6H_2O$$

the respiratory quotient is $\frac{6}{6}$, that is 1, and when triolein is oxidized according to the equation

$$C_{57}H_{104}O_6 + 80O_2 = 57CO_2 + 52H_2O$$

the R.Q. is $\frac{57}{80}$, or 0·71. It can also be shown that the respiratory quotient for an average protein is 0·80 (Table 11, 3). For a subject on an ordinary mixed diet the respiratory quotient represents the oxidation of a mixture of carbohydrate, fat and protein, and is of the order of 0·85. The respiratory quotient therefore gives us a clue to the type of food which is being oxidized by the body at any given time if we assume that the carbon dioxide collected and measured is derived solely from the oxygen taken into the body during the same period.

TABLE 11, 3

Metabolic 'Constants'

	Physiological fuel value kcal per g.	R.Q.	Calorific value of oxygen kcal./litre at s.t.p.d.
Carbohydrate	4	1·00	5·047
Fat	9	0·718	4·735
Protein	4	0·802	4·463

THE CALORIFIC VALUE OF OXYGEN

When carbohydrate is oxidized in the body to carbon dioxide and water for each litre of oxygen used up 5·047 kilocalories are produced; thus the calorific value of oxygen used in the combustion of starch is said to be 5·047 kilocalories per litre. In the same way for the combustion of human fat the calorific value of oxygen is 4·735 kilocalories per litre and for protein the value is 4·463 kilocalories per litre (Table 11, 3). It is therefore clear that the calorific value of a litre of oxygen used by the body varies between 4·5 and 5 approximately, depending on the types and proportions of the foodstuffs being metabolized, which will in turn be indicated by the respiratory quotient. The calorific values of oxygen used by the body in the combustion of mixtures of fat and carbohydrate for different levels of the respiratory quotient between 0·7 and 1·0 are shown in Fig. 11, 4.

As a first approximation (since the correction needed is small) we may neglect protein metabolism in estimating by indirect methods the heat output of the body. A measurement of the oxygen consumed and the carbon dioxide produced gives the respiratory quotient from which the calorific value of oxygen may be determined for the particular conditions under investigation. The volume of oxygen consumed during the period of the experiment multiplied by this value gives the heat output during the same period.

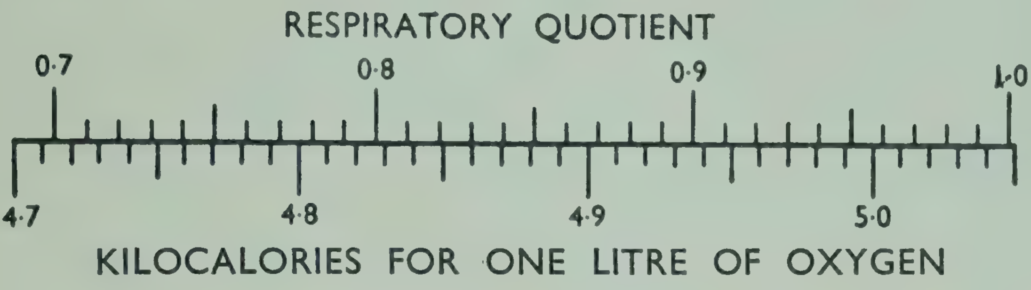


FIG. 11, 4

A figure (nomogram) from which may be read off the calorific value (C_{cf}) of a litre of oxygen (at s.t.p.) at non-protein R.Q.s (R_{cf}) from 0·7 to 1·0. (From G. H. Bell (1952), *Experimental Physiology*, p. 257. Glasgow: Smith & Son.)

The calorific value of a litre of oxygen (C_{cf}) at a non-protein respiratory quotient (R_{cf}) is given by the equation

$$C_{cf} = 3{\cdot}941 + 1{\cdot}106\ R_{cf} \quad . \quad . \quad . \quad . \qquad (1)$$

which is evaluated in the nomogram of Fig. 11, 4.

If carbohydrate alone is being metabolized $R_{cf} = 1$, so that

$$C_{cf} = 3{\cdot}941 + 1{\cdot}106 = 5{\cdot}047,$$

which is the calorific value of a litre of oxygen used in the oxidation of carbohydrate (Table 11, 3). If fat alone is being burned R_{cf} is 0·718, so that

$$C_{cf} = 3{\cdot}941 + (1{\cdot}106 \times 0{\cdot}718) = 4{\cdot}735,$$

which is the value in Table 11, 3.

In some forms of metabolic apparatus the volumes of oxygen used and of carbon dioxide produced are measured. In such a case the heat produced is given by the equation:

$$\begin{aligned}\text{Total kcal} &= (3{\cdot}941 + 1{\cdot}106\ R_{cf}) \times \text{ vol. of oxygen (in litres)}\\ &= \left(3{\cdot}941 + 1{\cdot}106\,\frac{\text{vol. of } CO_2}{\text{vol. of } O_2}\right) \times \text{vol. of } O_2\\ &= 3{\cdot}941 \times \text{vol. of } O_2 + 1{\cdot}106 \times \text{vol. of } CO_2.\end{aligned}$$

Now consider the effect of introducing protein. As the end-products of protein metabolism (mainly urea) are still theoretically capable of further combustion, the calorific value of a litre of oxygen metabolizing protein is less than that indicated by the bomb calorimeter. Its value, 4·463, is less than that for carbohydrate and fat metabolism at the same R.Q. and is in fact less than that for fat itself. If the effect of protein is neglected, as is often done, and if the total R.Q. is used instead of the non-protein R.Q., too high a value for the metabolic rate is obtained.

The classical method of overcoming the difficulty is to calculate from the urinary nitrogen the amount of protein metabolized in the body during the time of the experiment so that the volumes of oxygen used and of carbon dioxide produced in protein metabolism may be obtained. When these volumes are deducted from the corresponding total volumes, the volumes of oxygen and carbon dioxide used in carbohydrate-fat metabolism are found. Hence the non-protein R.Q. can be obtained. For this value of R.Q. the calorific value of a litre of oxygen is obtained from equation (1), or from the nomogram, and the heat derived from the metabolism of carbohydrate and fat can be calculated as before. Since we know from the urinary nitrogen the amount of protein metabolized we can calculate the number of kilocalories derived from the oxidation of protein alone, and this value added to the value obtained from non-protein metabolism gives the total number of kilocalories produced during the experimental period.

The following simpler method has been devised by Weir.

It can be shown that the value in kilocalories, C_{cpf}, of a litre of oxygen metabolizing carbohydrate, protein and fat is given by

$$C_{cpf} = 3{\cdot}941 + 1{\cdot}106R - 0{\cdot}365y \qquad . \quad . \quad . \qquad (2)$$

where R is the respiratory quotient and y the number of litres of oxygen metabolizing protein per litre total oxygen consumption.

On substituting the appropriate values from Table 11, 3 for R and y, it will be found that the formula gives the correct calorific value of oxygen for any one foodstuff, carbohydrate, protein or fat. For example, if protein alone is being metabolized, $R = 0{\cdot}802$, $y = 1$ and $C_{cpf} = 3{\cdot}941 + 1{\cdot}106 \times 0{\cdot}802 - 0{\cdot}365 = 4{\cdot}463$ which is the value in Table 11, 3.

From the formula it follows that the calorific value of a litre of oxygen metabolizing a mixture of carbohydrate, protein and fat can be obtained by deducting $0{\cdot}365y$ from the value found by using the total R.Q. instead of the non-protein R.Q. in the nomogram (Fig. 11, 4) or in equation (1). Over a period of time the deduction or correction for protein is $0{\cdot}365y$ for each litre of total oxygen consumption and, therefore, the total deduction is 0·365 times the number of litres of oxygen metabolizing protein during that time. Now it has been estimated that each gram of urinary nitrogen represents the consumption of 5·941 l. of oxygen in the metabolism of 6·25 g. of protein. The correction corresponding to 1 g. of urinary nitrogen is, therefore, $0{\cdot}365 \times 5{\cdot}941 = 2{\cdot}17$ kcal and the correction for 1 g. of protein metabolized is $\frac{2{\cdot}17}{6{\cdot}25} = 0{\cdot}347$ kcal.

In a given period of time, therefore, the heat produced is given by the equation:

$$\text{Total kcal} = (3{\cdot}941 + 1{\cdot}106\ R) \times \text{l.}\ O_2 \text{ used} - \text{protein correction}$$

or

$$3{\cdot}941 \times \text{l.}\ O_2 \text{ used} + 1{\cdot}106 \times \text{l.}\ CO_2 \text{ produced} - \text{protein correction}$$

and the protein correction is 2·17 × number of g. urinary nitrogen produced in that time or 0·347 × number of g. of protein metabolized in that time. For example if 62·5 g. protein are metabolized per day then 21·7 kcal must be deducted from the value obtained by using the formula $3{\cdot}941 + 1{\cdot}106\,R$, that is approximately 1 per cent. if that value were 2000 kcal.

It should be noted that the R.Q. may sometimes lie outside the limits 0·71 to 1·0. Thus when fat is being laid down by the transformation of carbohydrate, the non-protein R.Q. is greater than 1, since carbon dioxide is produced without oxygen being used up,

$$13C_6H_{12}O_6 \rightarrow C_{55}H_{104}O_6 + 23CO_2 + 26H_2O.$$

Conversely when fat is being transformed to carbohydrate in excess of that being oxidized, the non-protein R.Q. is less than 0·71 because oxygen is used up without a corresponding production of carbon dioxide. During gluconeogenesis (p. 296) figures less than 0·7 may be obtained.

The formulæ $C_{cf} = 3{\cdot}941 + 1{\cdot}106R_{cf}$ and $C_{cpf} = 3{\cdot}941 + 1{\cdot}106R - 0{\cdot}365y$ also hold when, instead of actual combustion of all the foodstuffs, there is a metabolic transformation from one to another, provided that there is no accumulation or loss of intermediate products.

METHODS FOR ESTIMATING ENERGY OUTPUT

Since the body obtains its energy by the combustion (oxidation) of food, and because this combustion involves the production of heat, the energy output of an individual may be assessed by measuring his heat production over a known period of time.

1. **Direct Calorimetry.**—For the determination of energy expenditure by direct calorimetry the subject is placed in a calorimeter consisting of a small room with heavily insulated walls (Fig. 11, 5). A continuous circulation of air is maintained through the chamber, oxygen being added to make good that which is used up during respiration. The carbon dioxide produced by the subject is absorbed by soda-lime. A porthole in the wall of the chamber is provided so that food may be passed in and excreta passed out; the subject can therefore live inside the calorimeter for several days. A current of water flowing through a system of pipes removes the heat produced by the subject, and by multiplying the difference in temperature between the ingoing and outgoing water by the volume of water flowing his heat output can be obtained. With this equipment it is possible to make very accurate measurements of the heat output either of the human subject or of experimental animals but, since it is elaborate and involves the use of complicated and very costly apparatus requiring continuous supervision, it is found only in a few laboratories.

2. **Indirect Calorimetry.**—Since the method of direct calorimetry is so complicated, indirect methods have been evolved in which the heat output of the subject is calculated from his oxygen consumption and his carbon dioxide output.

(A) *Closed Circuit Methods.*—In the most accurate type of closed circuit method the subject is kept in a respiration chamber (Fig. 11, 6) through which a continuous circulation of air is maintained by means of a rotary blower. Oxygen is added through a meter to make good the oxygen consumed in respiration; the water and carbon dioxide in the expired air are caught in sulphuric acid and soda-lime traps included in the ventilation circuit and are estimated by weighing or by a chemical method. The chamber is airtight but since it is not insulated the heat production cannot be obtained directly. The subject remains within the chamber for a period of several hours and his oxygen consumption and carbon dioxide output are measured. In this way his respiratory quotient may be calculated. Protein metabolism can be taken into account by measuring the nitrogen content of the urine excreted during the time of the experiment and a suitable correction may be made as described on p. 156.

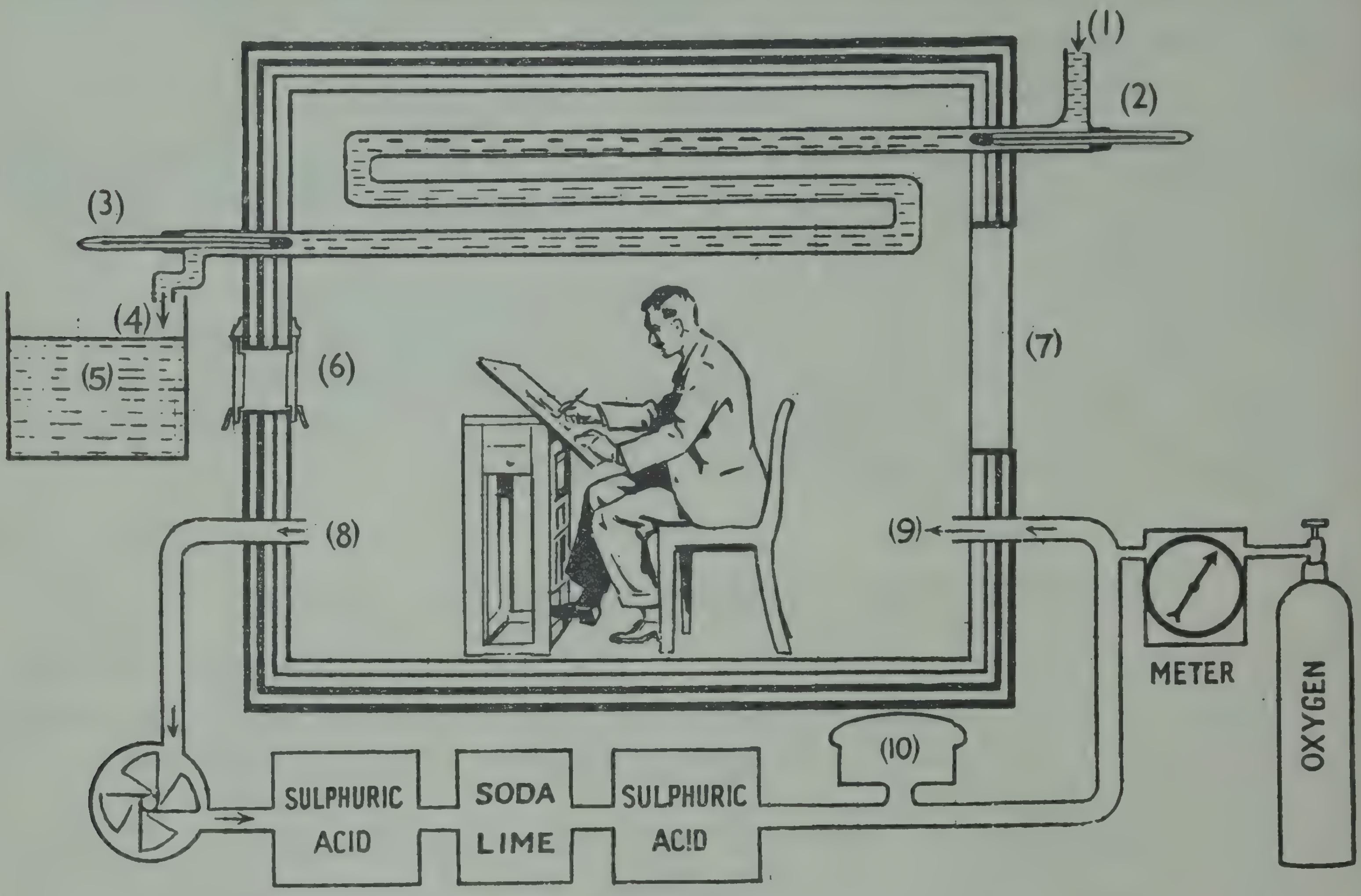


Fig. 11, 5

Atwater-Benedict Respiration Calorimeter

The walls of this chamber are insulated. Heat produced in it is absorbed by water passing in at (1) and out at (4), its temperature on entering and leaving being recorded on the thermometers (2) and (3). The volume of water which has flowed through the cooling system is measured in the vessel (5). The subject may be observed through the window (7), while food may be introduced and excreta removed through the porthole (6). Air leaves the chamber at (8) and passes through a blower and over sulphuric acid and soda-lime to absorb water and carbon dioxide. Oxygen measured by a gas meter is added to the system before the air passes into the chamber again at (9). (10) is a tension equalizer.

For clinical work, however, much simpler methods have been evolved. The most important of these is the recording spirometer of the Roth-Benedict type (Fig. 11, 7). This consists of a gasholder which is filled with oxygen from a cylinder. The top of the gasholder is connected by a cord running from a pulley to a writing lever which records the level of the gasholder on a rotating drum. The holder is connected by a system of wide-bore rubber tubes to a mouthpiece through which the patient must breathe when his nose is closed by a clip. The patient breathes in oxygen from the gasholder and his expired air passes through a valve into a soda-lime container to absorb the carbon dioxide and so back into the gasholder. As oxygen is used up the level of the gasholder falls and from the record

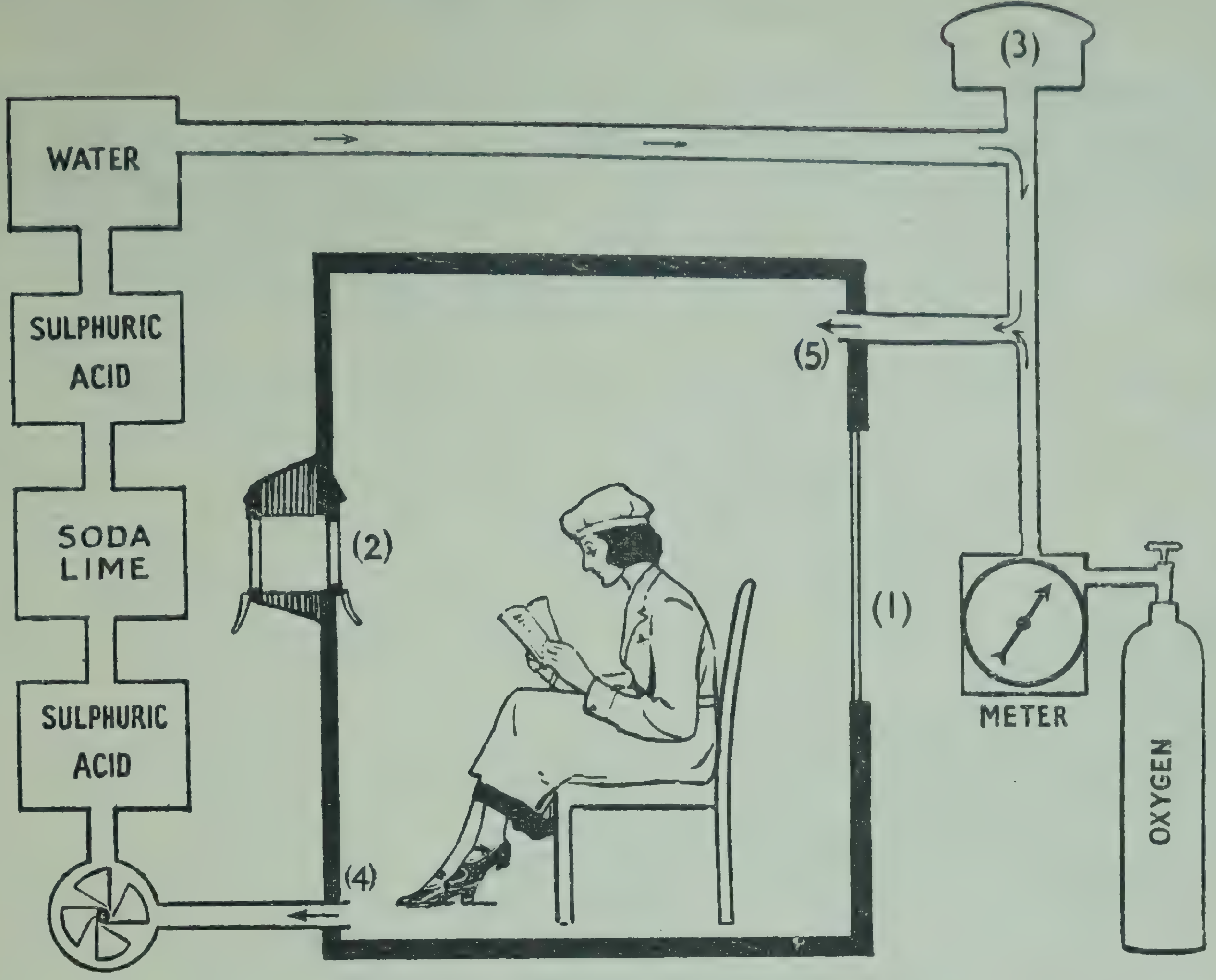

WATER
SULPHURIC ACID
SODA LIME
SULPHURIC ACID
(1)
(2)
(3)
(4)
(5)
METER
OXYGEN

Fig. 11, 6

Respiration Chamber

The chamber is provided with an observation window (1) and a porthole (2), for passing in food and removing excreta. The air in the chamber passes out at (4), through the retaining blower and over sulphuric acid and soda-lime to absorb water and carbon dioxide. It is then passed over water to restore humidity and re-enters the chamber at (5), tension being equalized at (3). Oxygen enters the circuit through the meter.

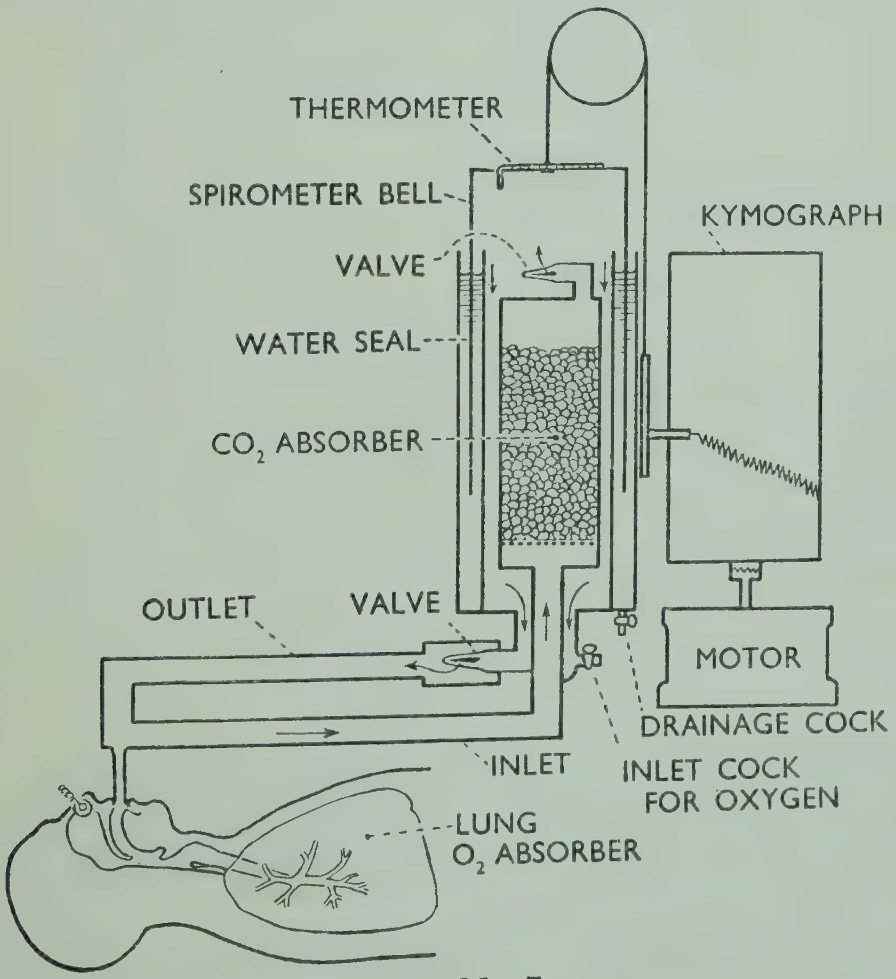

THERMOMETER
SPIROMETER BELL
VALVE
WATER SEAL
CO₂ ABSORBER
KYMOGRAPH
OUTLET
VALVE
MOTOR
DRAINAGE COCK
INLET
INLET COCK FOR OXYGEN
LUNG
O₂ ABSORBER

Fig. 11, 7

Diagram of Recording Spirometer

The subject wears a nose-clip and breathes through a mouthpiece which is connected to the apparatus by two tubes. He breathes in oxygen through the inspiratory valve and breathes out through the expiratory valve into the carbon dioxide absorber. The amount of oxygen used is recorded on the revolving drum by the pen attached to the counter weight.

(Diagram by J. B. de V. Weir.)

of its movements on a rotating drum the amount of oxygen consumed in a given period can be measured (Fig. 11, 8). Since the carbon dioxide is not measured, the respiratory quotient cannot be ascertained but an arbitrary value of 0·8 is usually taken. For this respiratory quotient the calorific value of oxygen is 4·825 kilocalories per litre, and by multiplying this value by the number of litres of oxygen consumed the energy output during the experimental period is obtained with an error of no clinical importance.

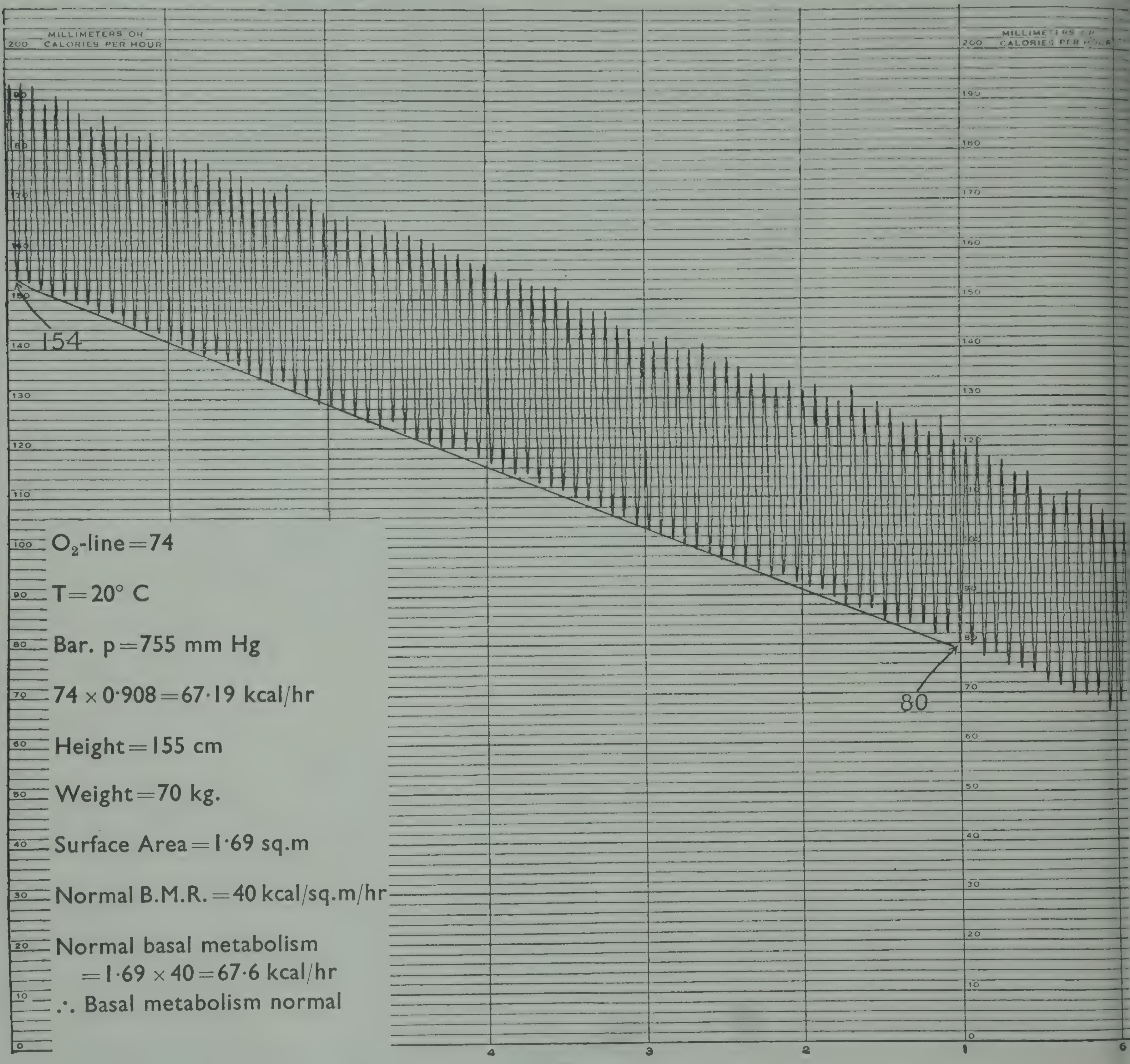


Fig. 11, 8

Chart of oyxgen consumption and data for the estimation of B.M.R., which is expressed in kcal per sq.m per hour. The calculations are explained in the text. (*After* E. W. H. Cruickshank (1951), *Food and Nutrition*. Edinburgh: Livingstone.)

The following example is taken from Cruickshank's *Food and Nutrition*. The subject of the experiment, a man aged twenty-five, weighed 70 kg. (154 lb.), and his height was 155 cm (61 in.). From Fig. 11, 12 it will be seen that his surface area is 1·69 sq.m. Fig. 11, 8, shows his record on the Roth-Benedict spirometer. The straight line drawn along the lower edge of the trace shows that during a measured period of six minutes the spirometer bell fell through a distance of 154 − 80, that is 74 mm. The dimensions of the bell are such that a fall of 1 mm corresponds to an oxygen uptake of 20·73 ml. If the calorific value of oxygen is arbitrarily taken as 4·825 kcal per litre, then a fall of 1 mm corresponds to $\frac{20{\cdot}73 \times 4{\cdot}825}{1000}$, that is 0·100 kcal. A fall of 74 mm in six minutes there-

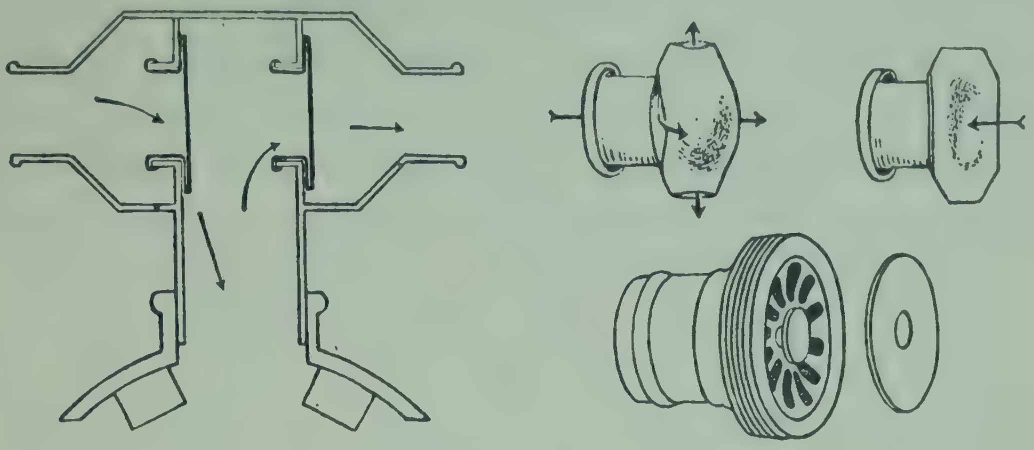


FIG. 11, 9

The diagram on the left shows the mouthpiece and valve housing and the path of the air to the Douglas bag which is connected by a wide-bore rubber pipe to the right-hand tube of the housing. The two small sketches on the right above indicate how the rubber valves behave according to the direction of the air stream. The sketch on the right below shows an alternative valve arrangement. A perforated circle of soft rubber is pushed over the 'collar stud' in the centre of the radially perforated plate. Air moving left to right meets very little obstruction but air cannot pass in the other direction as the rubber circle covers over the holes. (From G. H. Bell (1952), *Experimental Physiology*, p. 231. Glasgow: Smith & Son.)

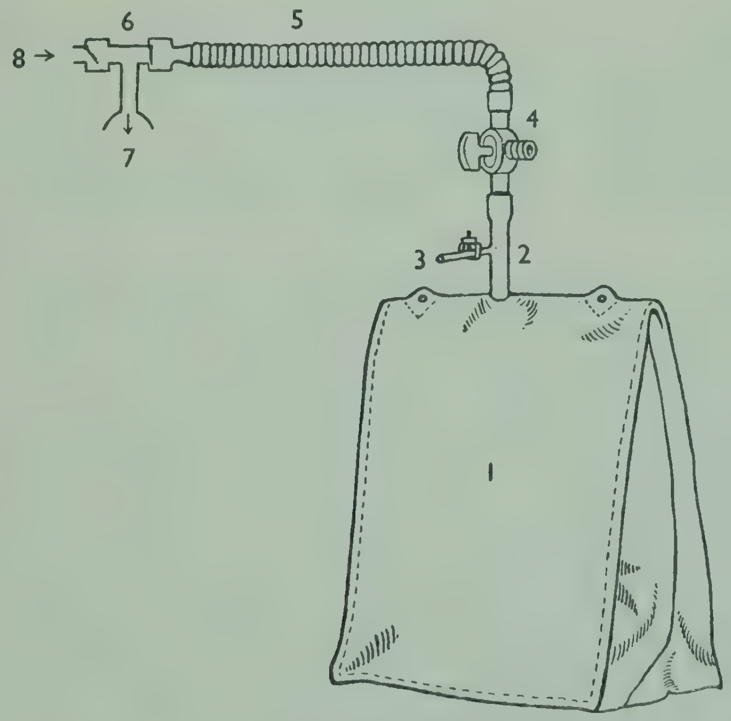


FIG. 11, 10

The Douglas Bag

(1) Douglas bag, capacity 100 litres. (2) Neck of bag. (3) Side tube for sampling, normally closed by screw-clip. (4) Tap for closing bag. (5) Hosepipe connecting bag with mouthpiece (7) by way of valves (6). Air to be inspired enters at (8). Expired air passes by way of the valves (6) and the hosepipe (5) into the bag (1).

fore indicates 74 × 0·1 × 10, that is 74 kcal per hour. A correction must however be made for temperature, water vapour and barometric pressure. If the temperature is 20° C and the pressure is 755 mm Hg the correction factor is 0·908 and the energy output is 74 × 0·908, or 67·19 kcal per hour.

The normal value for a man of 25 with a surface area of 1·7 sq.m is 40 × 1·7, or 67 kcal per hour (Table 11, 16) under basal conditions.

(B) *Open Circuit Method.*—In the open circuit method the subject's nose is closed by a clip and he breathes through a mouthpiece fitted with valves (Fig. 11, 9) so that he inspires the ordinary air of the room, while his expired air is carried through a wide tube to a rubberized canvas bag or *Douglas bag* of capacity 100 or 200 litres (Fig. 11, 10). The subject breathes into the bag for a period of time which is accurately measured, and the total volume of air in the bag is subsequently measured by means of a gas-meter. The subject's expired air must be collected until the bag is well filled so that any dilution by residual air in the bag is made negligible. A sample is taken for analysis in the Haldane gas-analysis apparatus and its carbon dioxide and oxygen content are determined by absorption with suitable reagents such as potassium hydroxide and alkaline pyrogallol. The volume of oxygen used, the volume of carbon dioxide produced and the respiratory quotient are calculated; the calorific value of oxygen may then be obtained from Fig. 11, 4. A simple calculation gives the number of kcal produced during the period of the experiment. This method has the advantage that the apparatus is easily portable. It may, for example, be carried about the wards of a hospital, or the bag may be strapped to a subject's back and his energy output during exercise, such as running or cycling, may be determined.

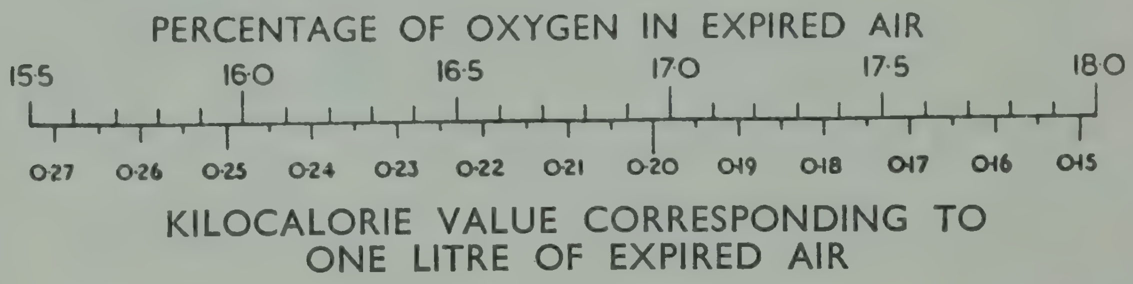


Fig. 11, 11

Calorific value (C_{air}) of one litre of expired air (neglecting protein correction) obtained by the Douglas bag technique. Although only the oxygen percentage is used this formula is accurate and allows for differences in calorific value with differences of R.Q. (J. B. de V. Weir (1949), *J. Physiol.* **109,** 1).

The calculation of the metabolic rate becomes very simple if, instead of calculating how much oxgyen has been used and multiplying it by the appropriate calorific value per litre as described above, the volume of expired air is multiplied by the calorific value corresponding to one litre of expired air.

It has been shown by Weir that this value C_{air} is given by the equation

$$C_{air} = 1{\cdot}048 - 0{\cdot}05O_E$$

where O_E is the percentage of oxygen in the expired air. The nomogram in Fig. 11, 11, may also be used. The corresponding equation including a 1 per cent. protein correction is

$$C^P_{air} = 1{\cdot}046 - 0{\cdot}05O_E$$

where C^{P}_{air} is the calorific value corresponding to a litre of air when the average protein correction is included. The following example may make the method clearer.

The subject of the experiment breathed into a Douglas bag for ten minutes and the volume of expired air as measured in a gas-meter was 99·75 litres at 16·8° C and 750 mm Hg atmospheric pressure. A sample of this air analysed in the Haldane apparatus gave 16·86 per cent. oxygen, 3·54 per cent. carbon dioxide, 79·60 per cent. nitrogen. The subject weighed 50 kg. and his height was 163 cm. The urinary nitrogen was 0·5 g. per hour. The volume of expired air corrected to s.t.p.d. was 99·75 × 0·912 = 90·97 litres. From Fig. 11, 11 it can be seen that the calorific value of one litre of expired air containing 16·86 per cent. oxygen is 0·205, the output is 90·97 × 0·205 = 18·65 kilocalories in ten minutes or 111·9 kilocalories per hour. The correction for protein metabolism (see p. 156) is 0·5 × 2·17 = 1·09 kilocalories per hour and the corrected output is therefore 111·9 − 1·1 = 110·8 kilocalories per hour. The surface area of a man of this size from the nomogram in Fig. 11, 12 is 1·52 sq.m and his metabolic rate is therefore $\frac{110 \cdot 8}{1 \cdot 52} = 72 \cdot 9$ kilocalories per sq.m per hour.

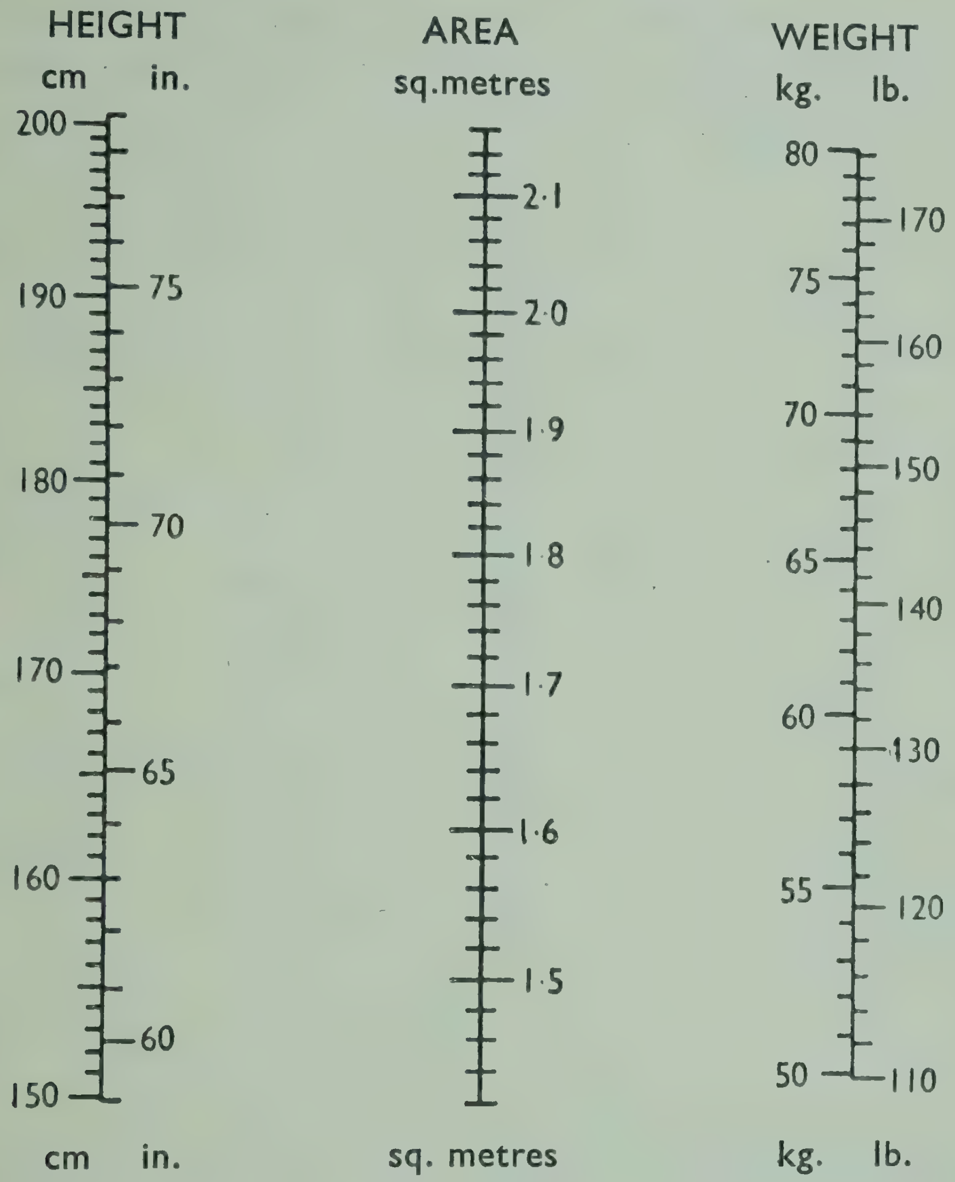


FIG. 11, 12

Nomogram for calculating surface area. A line joining the height on the left-hand scale to the nude weight on the right-hand scale cuts the centre scale at the surface area predicted by Du Bois's formula given on p. 167. (From G. H. Bell (1947), *Experimental Physiology*. Glasgow: Smith & Son.)

THE EFFICIENCY OF THE HUMAN BODY

A very important consideration in the working of an engine is its efficiency. This is given by the ratio of the useful work done to the total potential energy of the fuel supplied. The difference between the total potential energy and the useful work appears as heat. Thus

$$\text{Efficiency} = \frac{\text{useful work}}{\text{P.E. of fuel}} = \frac{\text{useful work}}{\text{useful work} + \text{heat}}.$$

Although the body works in quite a different way from a heat engine estimations of its efficiency can be made. Their interpretation, however, may be somewhat difficult.

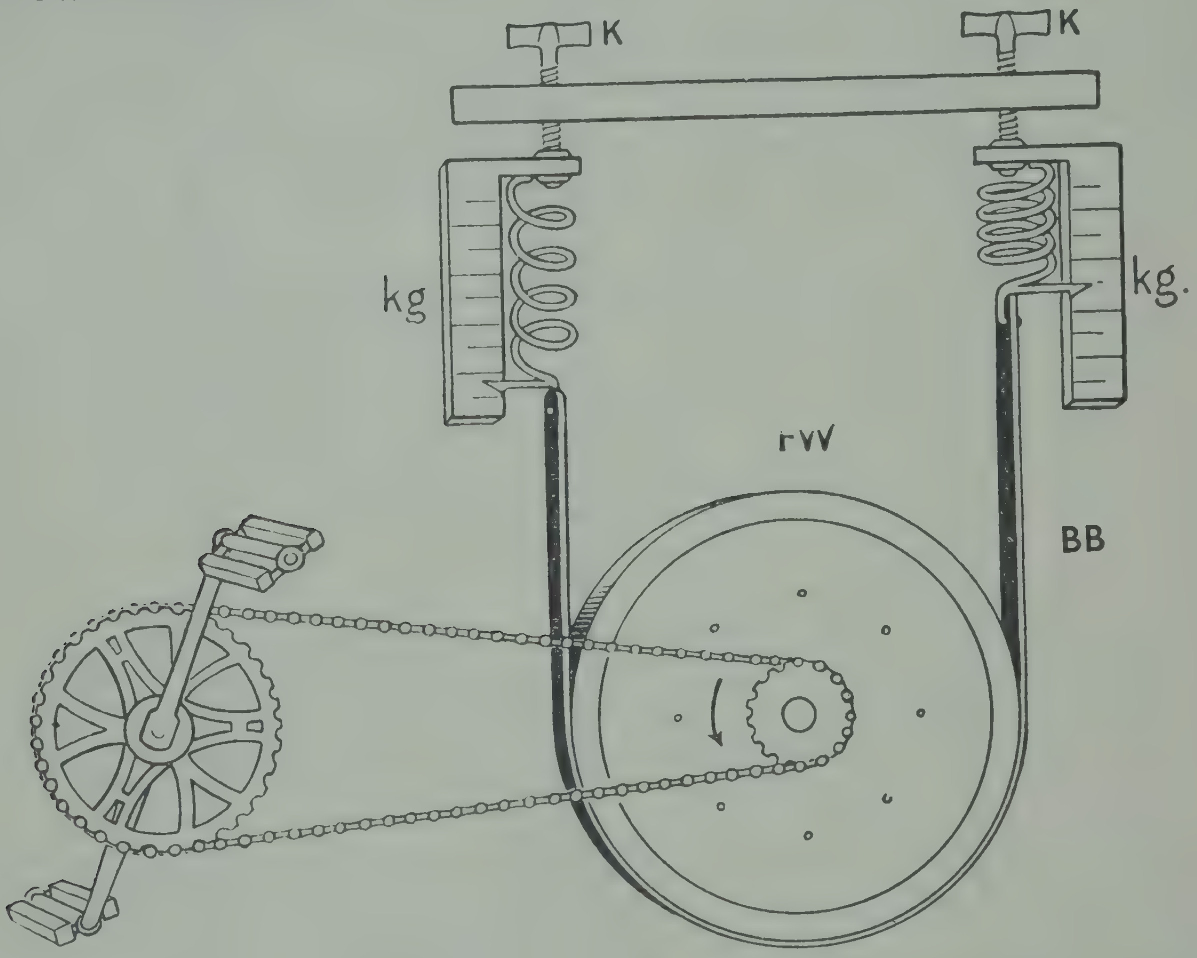


Fig. 11, 13

The subject sits on a fixed bicycle and drives a heavy flywheel (FW) by means of the usual bicycle pedals and chain. A brake band (BB) passed round the periphery of the flywheel is kept tense by two coiled springs. The force with which the springs are extended can be read on the kg. scale. The work done, kg.m per revolution of the flywheel, is equal to the difference between the readings on the two scales multiplied by the circumference of FW in metres. The work done per revolution can be adjusted by the keys (K).

In most of the experiments designed to test efficiency the subject has pedalled a stationary bicycle connected to an ergometer which may consist of a band brake or an eddy-current brake. This apparatus brings into action a very large mass of muscle tissue. The principle of the band brake is illustrated in Fig. 11, 13. The subject pedals and makes a flywheel revolve against a band brake which is kept pressed against it by the pull of two dynamometers (spring balances). The difference in reading in kg. of the two dynamometers multiplied by the circumference of the wheel in metres gives the work done in kilogram-metres (kg.m) per revolution. In a measured period of time the expired air of the subject can be collected and the total energy output of the body calculated by the methods already described; from the number of revolutions of the flywheel recorded during the same period the work done is calculated and converted to its heat equivalent (1 kcal = 427 kg.m). The ratio of these two values is usually called the gross efficiency. This value is however not of very great interest, since the body, unlike an engine, has to maintain itself even when it does no external work. At rest then the gross efficiency of the body is zero; the gross efficiency increases with the work done.

It is much more valuable to know the metabolic cost of the useful work done. This is called the net efficiency and is given by the equation

$$\text{Net efficiency} = \frac{\text{Heat equivalent of external work}}{\text{Increase in energy output}}.$$

The values for net efficiency are somewhat arbitrary because it is difficult to know what should be taken as the base line in an ergometer experiment. The base line might be the metabolism of the subject lying down, sitting on the bicycle without pedalling, or sitting on the bicycle with the pedals driven by an electric motor so that the subject's feet are carried round. In most experiments the base line is taken as the subject's metabolic rate lying down. Thus

$$\text{Net efficiency} = \frac{\text{Heat equivalent of external work}}{\begin{array}{l}\text{Metabolism}\\ \text{during the}\\ \text{working}\\ \text{period}\end{array} \quad \textit{less} \quad \begin{array}{l}\text{Metabolism}\\ \text{when lying}\\ \text{down for the}\\ \text{same time}\end{array}}.$$

Cathcart and Benedict found that in severe exercise the *gross* efficiency of untrained subjects was 10 to 20 per cent., whereas their trained subject, a professional cyclist, had a *gross* efficiency of 21 per cent. as a maximum. The *net* efficiency of the untrained subject was 21 to 23 per cent.; the *net* efficiency of the trained subject was about 25 per cent. They also found that there was an optimum rate of pedalling; if the rate were too high the efficiency fell off (see also Chapter 42).

It might be supposed that the most efficient person would carry out work at the lowest metabolic cost. This idea is however applicable to the human engine only up to a certain point, as will be seen by an experiment quoted by Cathcart and Benedict.

The efficiency of a highly trained subject doing most strenuous work on a bicycle ergometer was measured over a period of four hours. His efficiency at the start was 23·1 per cent. and in the observation made just before the experiment ended, due to the impending collapse of the subject, it was 21·3 per cent. One can state then, in this extreme example, that although the subject was near exhaustion the apparent efficiency gave no indication that shortly his effectivity would be nil.

It has been calculated that the efficiency of an average sprinter is about 23 per cent. He can incur an oxygen debt (see Chapter 42) equivalent to 13 horse-power while he does mechanical work—against gravity, in accelerating and decelerating the limbs—at the rate of 3 horse-power.

BASAL METABOLIC RATE

The output of energy by a living organism and the rate of oxygen consumption depend on the muscular activity, the nature and amount of the food eaten, changes in environmental and body temperature, emotional excitement and so on. In order to reduce the influence of these factors to a minimum, so that the metabolism of one individual may be compared with that of another, it is usual to make measurements of the metabolism of human beings under standard conditions. The subject is kept in bed, warm and comfortable, completely relaxed, mentally and physically, for 12 to 15 hours after the last meal. The expenditure of energy under these conditions is called the *basal metabolism*, which may be defined as the energy expenditure of the organism in a state of complete mental and physical rest. A certain amount of energy must continually be produced in order to maintain the body temperature at 37° and to allow for such essential processes as the beating of the heart and breathing. Under basal conditions the heat produced to replace that lost from the body forms a considerable part of the total, so that the basal metabolism is to a large extent dependent on the surface area of the body. The smaller the weight of an animal the greater is its surface area per unit weight; hence heat loss is greater per unit weight for a small individual than for a large one. This is illustrated by the following example.

TABLE 11, 14

Energy Expenditure by Different Species *

	Weight in kg.	kcal per day	
		Per kg.	Per sq.m surface
Pig	128·0	19·1	1078
Man	64·3	32·1	1042
Dog	15·2	51·5	1039
Goose	3·5	66·7	967
Fowl	2·0	71·0	943
Mouse	0·018	654·0	1188

* Taken mainly from E. Voit (1901), *Z. Biol.* **41,** 120.

A baby of one year weighs 10 kg. (22 lb.) and has a surface area of 0·50 sq.m. His surface area per unit weight is therefore 0·05 sq.m/kg. On the other hand a man of twenty-five years weighing 68 kg. (150 lb.) and having a surface area of 1·8 sq.m has a surface area of only 0·027 sq.m/kg., that is the child has nearly twice the surface area per unit weight. The basal requirements are however not so very different in the two cases—for the man 936 kcal/sq.m/day and for the child 1140 kcal/sq.m/day.

Both mouse and man have an energy expenditure in moderate activity of about 1000 kcal/sq.m/day although on a weight basis the energy expenditure of the mouse is about 20 times that of the man (Table 11, 14).

Since it is difficult to measure the area of an animal it is useful to know that the daily fasting heat production of adult mammals, including animals of such diverse size as mice and cattle, is on the average $70 \times (\text{body weight in kg.})^{3/4}$ kilocalories.

The *surface area* of the body may be determined directly by clothing the subject in thin, wax-impregnated garments which are subsequently removed, cut up and measured. This method, however, is obviously very tedious and not applicable to routine work. When the surface area for a number of persons has been found directly, it is possible to construct a formula which will take into account the varying sizes of individuals and will give the surface area with reasonable accuracy. The most commonly used formula is that of Du Bois,

$$S = W^{0{\cdot}425} H^{0{\cdot}725} \times {\cdot}007184,$$

where S is the surface area in sq. metres, W is the nude weight in kilograms, H is the height in centimetres. The most convenient method of evaluating this formula is by means of the nomogram shown in Fig. 11, 12. Height, weight and surface area of children and adolescents of both sexes are shown in Table 11, 15.

TABLE 11, 15

Height, Weight and Surface Area of Children and Adolescents *

Age, years	Males					Females				
	Height		Weight		Surface area in sq. metres	Height		Weight		Surface area in sq. metres
	in.	cm	lb.	kg.		in.	cm	lb.	kg.	
1	30	76	22	10	0·50	30	76	20	9	0·50
2	33	84	27	12	0·57	33	84	26	12	0·57
3	38	95	33	15	0·64	38	95	33	15	0·64
5	42	107	41	18	0·72	42	107	40	18	0·71
8	49	124	54	24	0·92	48	122	51	23	0·88
11	55	140	70	32	1·12	54	137	71	32	1·10
12	57	145	77	35	1·20	57	145	80	36	1·22
13	59	150	86	39	1·28	60	152	90	41	1·32
14	61	155	97	44	1·38	61	155	100	45	1·40
15	64	162	111	50	1·51	63	160	108	49	1·48
16	66	168	122	55	1·61	64	162	115	52	1·54
18	68	172	134	61	1·72	65	165	124	56	1·60
20	69	175	140	64	1·76	66	168	130	59	1·65
25	69	175	147	67	1·82	66	168	133	60	1·67

* From E. W. H. Cruickshank (1952). *Food and Nutrition.* Edinburgh: Livingstone.

It is customary to express the basal energy production as the amount of energy produced per unit of surface area per hour. This value is called the *basal metabolic rate* (B.M.R.) and it is relatively constant for people of the same age and sex. The B.M.R. is frequently measured for diagnostic purposes and the results are expressed as a percentage of the normal value (taken as 100 per cent.) of a subject of the same sex and age. Variations within the range 85 to 115 per cent. or, as it is usually expressed, minus 15 to plus 15 per cent. are accepted as being within normal limits. The basal metabolic rate of human subjects of both sexes and of different ages is shown in Table 11, 16, from which it is seen that for a man the average

B.M.R. is about 40 kcal per square metre per hour and for a woman about 36 kcal per square metre per hour.

The B.M.R. is relatively low in the newborn child but it rises quickly after birth, and at the age of one year it is 46 to 47 kcal per square metre per hour. In the second year the value reaches a maximum and gradually declines as age advances. From Table 11, 16 it will also be seen that although the basal metabolic rate is higher in young children than in adults, the total basal kcal per twenty-four hours reach a maximum for men at full maturity and for women at about the age of seventeen. The B.M.R. is raised in hyperthyroidism and is low in myxœdema (Chapter 51).

TABLE 11, 16

Comparison of Basal and Total Metabolic Rates of Males and Females

From E. W. H. Cruickshank (1952). *Food and Nutrition.* Edinburgh: Livingstone.

Age	B.M.R. * kcal per sq. metre per hour		Total Basal Metabolism † kcal per 24 hours		Total kcal per day (U.S.A.) ‡	
	Males	*Females*	*Males*	*Females*	*Males*	*Females*
Birth	30	30	288	288	440	440
1 year	55	52	660	624	1000	1000
2 years	57	53	780	725	1200	1200
3 ,,	55	52	845	798	—	—
5 ,,	53	52	915	886	1600	1600
8 ,,	51·8	47·0	1143	993	2000	2000
11 ,,	47·2	45·2	1268	1193	2500	2500
12 ,,	46·8	43·3	1347	1267	—	—
13 ,,	46·5	42·0	1428	1330	—	—
14 ,,	46·4	41·5	1537	1391	3200	2800
15 ,,	46·0	40·0	1667	1420	—	—
16 ,,	45·7	38·8	1764	1434	3500	2600
18 ,,	43·2	37·5	1783	1440	3800	2500
20 ,,	41·6	36·3	1756	1437	3500	2400
25 ,,	40·3	36·0	1760	1442	3000	2400
40 ,,	38·0	35·0	1641	1344	3000	2400

* A similar table with somewhat different values for British people has recently been published by Robertson and Reid (1952). See list of references.

† Total basal metabolism = B.M.R. × 24 × surface area in sq.m.

‡ Recommended by the National Research Council of Washington, U.S.A., in 1948. Daily energy requirements of persons of normal activity. For British recommendations see Table 12, 17.

For clinical purposes the B.M.R. is always determined by indirect methods. The most popular method is the recording spirometer (Roth-Benedict type), which gives a result very quickly and is simple to use. The Douglas bag method gives a more accurate result but it is more tedious to carry out and requires greater technical skill. The patient must be carefully prepared in order to ensure that he is really in a basal state, and it is best that he should become acquainted with the apparatus and procedure before the actual test is carried out. A determination is made in the morning as soon as possible after the patient has wakened. He is usually made to fast for twelve hours before the test; his temperature should be taken to verify that it is normal; he must be comfortable and warm, and completely relaxed mentally and physically.

Specific Dynamic Action

It is important to note that after food is absorbed the metabolic activities of the body are stimulated with the result that the heat production increases. Thus an individual whose metabolism under basal conditions is 1700 kilocalories per day actually requires more than 1700 kilocalories of food in order to maintain himself in equilibrium. The metabolism of 1700 kilocalories of food raises his energy output to about 1900 kilocalories. This increase in metabolism, termed the *specific dynamic action* of foods, varies according to the type of food taken; for proteins it is about 30 per cent. of the basal metabolism, for fat and carbohydrates 4 to 5 per cent., and for a mixed diet it amounts to about 10 per cent. Thus for every thousand kilocalories in the basal metabolism 1100 kilocalories of food should be taken.

THE CALORIE REQUIREMENTS OF MAN

It will be obvious from what has already been said that it is possible to calculate the caloric requirements of the human subject with reasonable accuracy. Energy must be obtained from the food for three purposes. In the first place it is required for basal metabolism, that is to maintain the fundamental bodily activities, the heart beat, breathing, muscle tone and also the body temperature. For a man with a basal metabolic rate of 40 kcal/sq.m/hr, and a surface area of 1·8 sq.m, the caloric output for basal purposes will be $40 \times 1{\cdot}8$, that is 72 kcal/hr or 72×24, that is 1728 kcal/day. If we allow 10 per cent. for specific dynamic action his caloric requirements to maintain life will be $1728 + 173$ kcal, that is, about 1900 kcal per day.

In the second place we must give additional food to cover the expenditure of energy in the simple activities of everyday life, such as sitting, standing, walking, dressing and undressing, and so on, and for this purpose an allowance of 360 kilocalories for the eight hours of the day which are occupied by these simple activities may be made.

In the third place an allowance has to be made for work carried out during the day and this allowance must obviously depend upon the nature of the occupation. Thus people engaged in sedentary occupations expend 20 to 45 kilocalories per hour at their work, whereas men engaged in hard muscular work may expend as much as 400 kilocalories per hour. The energy expenditure per hour for people in different occupations is shown in Table 11, 17.

It is obvious then that, in order to calculate the total caloric requirement during twenty-four hours, a suitable allowance must be made for

TABLE 11, 17

From Ministry of Food, "Manual of Nutrition," 1947

	kcal/hr
(1) *Sedentary occupations*	
Writing	20
Typing	30
Tailoring	45
(2) *Moderatively active occupations*	
Shoemaking	90
Carpentering	140
Light-metal working	140
(3) *Active occupations*	
Heavy carpentering	180
Light blacksmithing	275
Stone-mason's work	300
(4) *Very active occupations*	
Coal-mining	320
Heavy blacksmithing	350
Wood cutting	380

NOTE.—It is important to remember that all these figures are averages. Two important factors which may influence them are (*a*) trained people can do work with less expenditure of energy than untrained people; (*b*) work of nominally the same character, for example labouring, may differ widely under different circumstances.

occupation. This has been done in Table 11, 18, where the total caloric requirements of four different people have been worked out. The total energy expenditure varies from 2420 to 5300 kilocalories per day. In order to meet this expenditure an equivalent amount of food must be consumed or else the body loses weight. Dietary surveys show that the caloric intakes of different people vary from 2000 to over 5000 kilocalories per day according to the type of work which they do.

The energy requirements of men, women and children have been assessed by the Committee on Nutrition of the British Medical Association whose figures are shown in Table 12, 17. They indicate that an average man requires 2250 to 5000 kcal per day in his food according to his occupation, while the average woman requires 2000 to 3750 kcal per day. During pregnancy and lactation the intake should be increased.

For children the energy requirements relative to the body weight are high, since children expend much energy in everyday activities, and in addition they require food for growth. Thus a girl of fourteen requires as much food in terms of kilocalories as does a woman engaged in active physical work. In the same way a boy of sixteen requires more food than does his father, unless the later is engaged in severe physical work.

These requirements are of course made up from the fats, proteins and carbohydrates of the diet, and of these carbohydrates provide by far the greater part of the total energy requirement. Thus a diet containing 400 g. of carbohydrate, 100 g. of fat and 100 g. of protein would provide 2900 kcal. The large proportion of carbohydrate in the diet is due simply to the relative abundance and cheapness of this component of the diet.

From the purely energy point of view the total requirements could be met by a wide variety of allocations of the caloric requirements among the three proximate principles. Although it is in fact possible to replace a given amount of one proximate principle by an isocaloric amount of another, such a process is limited by the considerations mentioned on p. 188 and p. 194.

	g.	kcal/g.	Total
Proteins . .	100	4	400
Fats . . .	100	9	900
Carbohydrates .	400	4	1600
			2900

TABLE 11, 18

The Day's Need for Energy *

The total daily caloric needs of men doing work of varying degrees of activity can be calculated as follows:

	kcal
(*a*) A clerk: sedentary occupation	
24 hours of the processes of living	1900
8 hours of everyday activity	360
8 hours of writing: 8 × 20	160
Total day's need	2420
(*b*) A carpenter: moderately active occupation	
24 hours of the processes of living	1900
8 hours of everyday activity	360
8 hours of carpentering: 8 × 140	1120
Total day's need	3380
(*c*) A blacksmith: active occupation	
24 hours of the processes of living	1900
8 hours of everyday activity	360
8 hours of blacksmithing: 8 × 275	2200
Total day's need	4460
(*d*) A woodcutter: very active occupation	
24 hours of the processes of living	1900
8 hours of everyday activity	360
8 hours of woodcutting: 8 × 380	3040
Total day's need	5300

---

* Modified from Ministry of Food, *Manual of Nutrition*, 1947.

It has been found that the average man doing physical work for eight hours a day is not likely to show evidence of fatigue if the intensity of the work and the length of the rest pauses are adjusted to give an average of less than 4 kcal per minute (in addition to the basal metabolism). This is equivalent to walking on the level at 3·8 miles per hour. If he works twice as fast he must have rest pauses equal to the working time.

Measurements of the height and weight of very large numbers of healthy people have been made and from this information it is possible to compute an "ideal" weight in relation to height, age and sex. A table of "ideal" weights is to be found in the appendix. Such figures provide a standard by which a given individual may be judged to be over or under weight. It is customary, however, to allow a variation of 15 per cent. above and below the "ideal" weight before regarding a person as abnormal.

REFERENCES

Benedict, F. G. & Cathcart, E. P. (1913). *Muscular Work.* Carnegie Institute Publications No. 187.

Cathcart, E. P. (1922). The efficiency of man and the factors which influence it. *Rep. Brit. Ass.* pp. 164–178.

Cruickshank, E. W. H. (1951). *Food and Nutrition*, 2nd ed. Edinburgh: Livingstone.

Garry, R. C., Passmore, R., Warnock, G. M. & Durnin, J. V. G. A. (1955). Studies on expenditure of energy and consumption of food by miners and clerks, Fife, Scotland, 1952. *Spec. Rep. Ser. med. Res. Coun., Lond., No.* 289. London: H.M.S.O.

Keys, A. & Brožek, J. (1953). Body fat in adult man. *Physiol. Rev.* **33,** 245–325.

Kleiber, M. (1947). Body size and metabolic rate. *Physiol. Rev.* **27,** 511–541.

Lusk, G. (1931). *The Elements of the Science of Nutrition*, 4th ed. Philadelphia: Saunders.

Mackworth, N. H. (1950). Researches on the measurement of human performance. *Spec. Rep. Ser. med. Res. Coun., Lond., No.* 268. London: H.M.S.O.

Manual of Nutrition (1950). London: Ministry of Food.

McCance, R. A. (1953). Overnutrition and undernutrition. *Lancet*, ii, 685–690 and 739–745.

Passmore, R. (1954). Recent investigations on human calorie requirement. *Chem. & Ind.* (*Rev.*), pp. 556–560.

Passmore, R. & Durnin, J. V. G. A. (1955). Human energy expenditure. *Physiol. Rev.* **35,** 801–840.

Robertson, J. D. & Reid, D. D. (1952). Standards for the basal metabolism of normal people in Britain. *Lancet*, i, 940–943.

Rynearson E. H. & Gastineau, C. F. (1949). *Obesity.* Oxford: Blackwell.

Sinclair, H. M. (1953). Assessment and results of obesity. *Brit. med. J.* ii, 1404–1407.

Symposium on energy expenditure in man (1956). *Proc. Nutr. Soc.* **15,** 72–99.

Weir, J. B. de V. (1949). New methods for calculating metabolic rate with special reference to protein metabolism. *J. Physiol.* **109,** 1–9.

Widdowson, E. M. (1955). Assessment of the energy value of human foods. *Proc. Nutr. Soc.* **14,** 142–154.

CHAPTER 12

FOOD AND NUTRITION

THE COMPOSITION OF FOODS

We have seen already how animal and plant tissues are made up of carbohydrates, fats and proteins together with water, minerals and vitamins. Since these tissues constitute a large part of the food we eat, it is necessary to study the composition of the commoner foodstuffs in terms of the relative amounts of carbohydrate, fat and protein which they contain. Analytical figures for some commoner foods are shown in Tables 12, 1, to 12, 10. As the composition of any one foodstuff may vary within wide limits, it must be emphasized that the figures in the tables are average values only. Certain of the important staple foodstuffs used by man will now be considered in greater detail.

Table 12, 1

Comparison of Human Milk with Cow's Milk

	Human g. per 100 g.	Cow's g. per 100 g.
Water	88	88
Protein	1·2	3·3
Fat	3·5	3·6
Carbohydrate	6·5	4·7
Ash	0·2	0·12
Caseinogen: lactalbumin ratio	1:2	3:1
Ascorbic acid	4 to 8 mg.	0·7 to 3 mg.

The composition of cow's milk varies with the breed of cow, with the time of lactation and with the nutritional state of the animal.

Milk.—The young mammal, including the human infant, starts life on a diet consisting solely of milk. It is therefore appropriate to consider first the chemical composition of milk, which contains all the important nutrients (Tables 12, 1, and 12, 2). The secretion of milk will be discussed in Chapter 52.

The chief proteins are caseinogen and lactalbumin. Caseinogen is a phosphoprotein present in milk as calcium caseinogenate. When acid is added to milk, caseinogen is rendered insoluble (its isoelectric point is pH 4·6) and the milk *curdles*. Lactic acid produced by bacteria when milk goes sour has a similar effect. On the other hand, when milk comes into contact with the enzymes rennin or pepsin, a process known as *coagulation* or *clotting* occurs in which the caseinogen is converted into the protein *casein* which forms an insoluble calcium salt (calcium caseinate). On standing, the clot contracts, expressing a clear fluid known as *whey*. The clot retains all the milk fat which contains a high proportion of lower fatty acids such as butyric acid.

The carbohydrate of milk is the sugar lactose.

The most important mineral elements in milk are calcium and phosphorus. Milk is deficient in iron, but the young mammal is born with a store of iron sufficient to tide it over the period of breast-feeding.

Milk is an important source of vitamins A and D, the content of which varies throughout the year according to the diet of the cow. It is a poor source of ascorbic acid.

TABLE 12, 2

The Composition of Cow's Milk *

Content per 100 g.

(1 pint is approximately 570 g.)

Calories	66	Carbohydrate	4·7 g.
Total solids	12·4 g.	Calcium	0·12 g.
Water	87·6 g.	Phosphorus	0·10 g.
Protein	3·3 g.	Magnesium	0·01 g.
Fat	3·6 g.	Iron	0·03 mg.
Vitamin A Carotene	150 i.u.	Pyridoxine Biotin	70 µg. 2·5 µg.
Thiamine	45 µg.	Inositol	18 mg.
Riboflavin	150 µg.	Vitamin C	2 mg.
Nicotinic acid	80 µg.	Vitamin D	2 i.u.
Pantothenic acid	250 µg.		

* S. K. Kon (1947), *Brit. med. Bull.* **5**, 170.

The composition of milk varies from one species to another and human milk differs considerably from cow's milk (Table 12, 1). The latter contains more protein and less carbohydrate than human milk and about the same amount of fat. Of the two chief proteins, lactalbumin is present in about the same concentration in both, but cow's milk contains three times as much caseinogen as lactalbumin, and human milk half as much. The amount of protein in the milk of any species varies with the rate of growth of the young of that species (Table 12, 3). Since the calf grows relatively more rapidly than the human infant it is not surprising to find that cow's milk has a higher protein content than human milk.

Many attempts have been made to make cow's milk more suitable for babies by altering its composition so that it resembles human milk more closely. The protein content is reduced to the level found in human milk by dilution, and cream and lactose may then be added to bring the fat and carbohydrate levels to the correct values. The resulting product is, however, still very different from human milk because dilution has not altered the casein-lactalbumin ratio or the amino-acid content of the proteins.

Milk is liable to be infected with pathogenic organisms either from the udder or from the persons handling it. The most important organism which may be introduced by the cow is the bacillus of bovine tuberculosis, and it has been estimated that some 5 to 10 per cent. of the farms in this country send out milk containing tubercle bacilli. Milk from herds which have been certified as free from tuberculosis is known as tuberculin-tested (T.T.) milk. It is not in the meantime possible to establish sufficient T.T. herds to supply milk for the whole population, but steps can be taken to render the milk free from this danger by the process of *pasteurization*. In the 'holder process' this consists in heating the milk rapidly to a temperature of 145° F (62·5° C), maintaining it at this temperature for thirty minutes and cooling quickly. A modification known as the 'high-temperature short-time' method is sometimes employed. Both processes destroy pathogenic organisms such as the tubercle bacillus or the bacillus of undulant fever which may come from the cow, and the organisms of diphtheria, typhoid fever, or scarlet fever which may be introduced during collection and distribution. Pasteurization cannot be relied upon to kill all the harmless bacteria in milk and therefore pasteurized milk can still become sour on keeping. Pasteurization has no material effect on the nutritive value of milk beyond a slight reduction in the content of thiamine (10 per cent.) and ascorbic acid (20 per cent.), but milk is not in any case an important source of either of these substances. The process does, however, destroy the enzymes of milk, including the enzyme *phosphatase*, and this is the basis of the delicate *phosphatase test* commonly employed to check the efficiency of pasteurization.

TABLE 12, 3

Comparison of the Milks of Different Species

	Mean values for protein content per cent.	Caseinogen: lactalbumin ratio	Time required to double birth weight (days)
Human . . .	1·2	0·5	180
Mare . . .	2·4	. .	60
Cow . . .	3·3	3	47
Goat . . .	4·1	. .	19
Pig . . .	5·9	. .	18
Sheep . . .	6·5	. .	10
Dog . . .	7·1	. .	8
Cat . . .	9·5	. .	7
Rabbit . . .	10·4	. .	6
Rat . . .	11·8	10	4·5

More drastic treatment, heating to 212° F for varying lengths of time, is used in the production of *sterilized milk* in which all bacteria, including the souring bacteria, are killed. This product has better keeping qualities than pasteurized milk but the taste is altered, the biological value of its proteins is slightly reduced and about half of the ascorbic acid content and one-third of the thiamine are destroyed.

Dried Milk.—In the preparation of dried milk, water is removed either by spray-drying or by roller-drying. The resulting product is of high nutritive value but cannot be stored for very long periods since the fat tends to be oxidized and become rancid. Dried skim milk is a valuable source of protein and calcium but is, of course, poor in fat and the fat-soluble vitamins A and D.

TABLE 12, 4

Composition of Dairy Produce

The figures in this and subsequent tables (12, 5, 6, 9 and 10) are taken from *Nutritive Values of Wartime Foods*, 1945 (tables compiled for the Accessory Food Factors Committee of the Medical Research Council) and refer to the "edible portion" of the food.

	Per 100 g.						
	Water g.	Protein g.	Fat g.	Carbohydrate g.	kcal	Calcium mg.	Iron mg.
Cow's milk, whole	88	3·3	3·6	4·4	63	120	0·1
Cow's milk, skimmed	91	3·4	0·1	4·6	33	124	0·1
Butter . . .	15	0·5	82·5	0·0	745	15	0·1
Margarine, vitaminized . .	14	0·0	85·3	0·0	768	4	0·3
Cheese, Cheddar .	37	24·9	34·5	0·0	410	810	0·6
Egg, whole, fresh .	74	12·5	11·5	0·9	157	60	3·0

Condensed Milk is prepared by removing part of the water *in vacuo*. The concentrate is then sterilized at a high temperature in tins, or it is treated with sufficient sugar (about 40 per cent.) to prevent the growth of bacteria.

In the preparation of *cream* the fat is separated from milk either mechanically or by allowing it to rise to the surface. Cream contains a high proportion (40 to 50 per cent.) of fat. Milk from which most of the fat has been removed is known as skim milk.

When milk fat (cream) separated in this way is submitted to prolonged shaking (churning) the fat globules coalesce to form a solid mass of *butter* which is therefore, like cream, an important source of the fat-soluble vitamins, especially vitamin A. Its natural colour is due mainly to carotene. The vitamin A and D content of butter is higher in summer, when the cows are fed on grass, than in winter. The residual fluid known as buttermilk is a rich source of protein, lactose and the inorganic salts of milk.

Cheese is made by coagulating the proteins of milk with rennet or with vinegar. Some of the fat is included in the coagulated mass which is pressed out and allowed to ripen under the influence of bacteria. The type of organism employed in this process is responsible for the characteristic flavour of any particular kind of cheese. Cheese is an important source of protein, fat, and mineral elements such as calcium, and has a very high nutritive value.

Margarine is prepared from unsaturated oils which are hydrogenated to yield a solid fat with a chemical composition very similar to that of butter. It contains a smaller proportion of the lower fatty acids and has a slightly different flavour. As prepared in this way margarine contains no vitamins, but vitamins A and D are now added in the proportions of 760 to 940 i.u. vitamin A per oz. and 80 to 100 i.u. vitamin D per oz. (28 g.).

TABLE 12, 5

Composition of Meat and Fish (*edible portion*)

	Per 100 g.						
	Water g.	Protein g.	Fat g.	Carbohydrate g.	kcal	Calcium mg.	Iron mg.
Beef, sirloin, for roasting . .	59	16·0	23·0	0·0	271	10	4·0
Beef, steak, for stewing . .	65	17·0	16·0	0·0	212	10	4·0
Mutton, loin and chops, for roasting	56	14·0	28·0	0·0	308	10	2·0
Bacon, sides as a whole . .	40	11·0	45·0	0·0	449	10	1·0
Tongue, canned .	57	19·0	20·0	0·0	256	10	3·0
Liver, ox . .	70	17·0	6·0	5·0	142	10	13·9
Cod or haddock, fresh . .	81	16·0	0·5	0·0	69	25	1·0
Sole or turbot, fresh	79	16·0	2·0	0·0	82	25	1·0
Herring, Jan.–Aug.	66	16·0	15·0	0·0	199	100	1·5
Herring, Sept.–Dec.	72	16·0	8·0	0·0	136	100	1·5
Salmon, fresh. .	67	17·0	14·0	0·0	194	25	1·0
Salmon, canned, bones eaten .	67	20·0	10·0	0·0	170	300	1·5
Haddock, smoked .	76	18·0	0·6	0·0	77	30	1·0
Kipper . .	62	19·0	16·0	0·0	220	120	2·0

Eggs are a rich source of protein, vitamins and minerals. One egg supplies 6·8 g. protein, 5·5 g. fat, 30 mg. Ca, 110 mg. P, 1·6 mg. Fe (all of which is available), 200–800 i.u. vitamin A, 0·1 mg. thiamine, 100–350 mg. riboflavin, 10–50 i.u. vitamin D and 80 kcal (Kon).

Meat.—Meat is essentially skeletal muscle. Analytical figures for some meats are shown in Table 12, 5. The values vary fairly widely according to the cut and the amount of fat present. When meat is cooked the proteins are coagulated, some water is lost and collagen is converted to gelatin. If the meat is boiled the soluble constituents such as inorganic salts, gelatin and extractives are lost, but these substances tend to be retained when the meat is roasted or fried. When fresh meat is allowed to hang carbohydrate (glycogen) disappears and acids such as lactic acid are produced which tend to soften the muscle fibres and make the meat more tender. In the case of game this process is allowed to proceed to the stage of incipient putrefaction.

The extractives in meat are soluble organic substances; the composition of many of them is not yet known. They are of little nutritive value but they are largely responsible for the flavour of meat and may serve to stimulate the flow of digestive juices. Concentrated preparations of these extractives are sold as meat extracts.

In addition to muscle other parts of animals usually described as offal are used for food. Liver and kidney are both rich in nucleoprotein and contain less fat than most meat. Liver is rich in vitamin A and in iron. Heart is similar in composition to skeletal muscle. Sweetbreads (pancreas and thymus) are rich in nucleoprotein and tripe (boiled ox stomach and intestine) in gelatin. Tripe also contains a high proportion of calcium which has been added in the form of lime during preparation. Brain has a low nutritive value.

Fish form an important source of animal protein (Table 12, 5). White fish such as cod, haddock and sole contain only a small proportion (less than 2 per cent.) of fat, but the fat fish, herring, mackerel and salmon, contain 5 to 18 per cent. of fat and can provide important amounts of the fat-soluble vitamins A and D. Large amounts of calcium and phosphorus may be added to the diet in sardines, sprats and canned salmon if the bones are eaten.

Cereals.—The crude cereal contains approximately 70 per cent. of carbohydrate in the form of starch granules with a thin indigestible cellulose coat, 11 per cent. of protein mainly glutelins and gliadins and fat; the amount of fat varies widely (0·5 to 8 per cent.) from one cereal to another, being particularly high in oatmeal. The inorganic matter, about 2 per cent., consists chiefly of calcium, phosphorus and iron, all of which are not however available for absorption. A large proportion of the phosphorus is in the form of phytic acid (inositol hexaphosphate). Phytic acid is not absorbed from the intestine and moreover it tends to combine with calcium and iron in the gut and so hinder their absorption. Consequently calcium and iron from other foods may be rendered unavailable if the diet contains much cereal. Finally most unprocessed cereals are good sources of the vitamins of the B complex and of vitamin E.

The various nutrients are not evenly distributed throughout the grain but tend to be localized in different regions. A section of the wheat grain is shown in Fig. 12, 7 and the proportions by weight of the various parts are given in Table 12, 8. The central core of the grain is the *endosperm*, consisting mainly of starch with some protein. The outer layer of the endosperm is rich in protein, minerals and nicotinic acid (a member of the vitamin B group). These substances also abound in the *aleurone layer* which lies immediately outside the endosperm. Outside this lie the very fibrous layers of the pericarp and the *testa*. At one pole of the grain lies the *germ*, consisting of the *embryo* bounded by the *scutellum*. The germ is particularly rich in certain vitamins (Table 12, 8).

The grain is crushed in the process of milling and the indigestible fibrous coats of the pericarp and testa are removed as bran, together with the aleurone layer. In the old process of stone-grinding the germ was retained, but in the modern process of roller-milling it is removed along with the fibrous layers. The resulting white flour has the advantage of good keeping qualities but much of the minerals, vitamins and proteins have been removed with the bran and embryo. When white bread was first introduced, in the second half of the nineteenth century, the intake of minerals and vitamins of the B group by the population of Great Britain was much reduced.

TABLE 12, 6

Composition of Cereal Products

	Per 100 g.						
	Water g.	Protein g.	Fat g.	Carbohydrate g.	kcal	Calcium mg.	Iron. mg.
Bread, brown, unfortified . .	37	8·5	1·3	50·4	247	18	2·1
Bread, made from 85 per cent. extraction U.K. flour fortified with Ca. . . .	37	8·5	1·2	51·4	250	57	1·8
Flour, Manitoba, 85 per cent. extraction fortified with Ca. . . .	13	13·9	1·7	69·0	347	81	2·9
Flour, Manitoba, 70 per cent. extraction unfortified .	13	13·1	1·2	71·6	350	13	1·8
Oatmeal . .	9	12·1	8·7	65·5	389	53	4·1
Pearl barley . .	11	7·7	1·7	73·2	339	10	0·7
Bran, wheat, processed . .	8	10·9	4·2	54·5	299	98	12·9
Germ, wheat, processed . .	8	32·0	7·7	37·8	349	58	9·7
Biscuits, plain (water) . .	4	12·0	3·0	75·0	375	80	2·1
Cakes and buns (plain)	30	8·0	7·0	48·0	287	50	1·0
Rice, polished, raw.	11·7	6·8	1·0	86·8	393	3·7	0·5

'Brown bread' is usually made from white flour to which a proportion of bran has been added. It has a higher vitamin content than has white bread, has more flavour and contains more indigestible fibrous matter or 'roughage.'

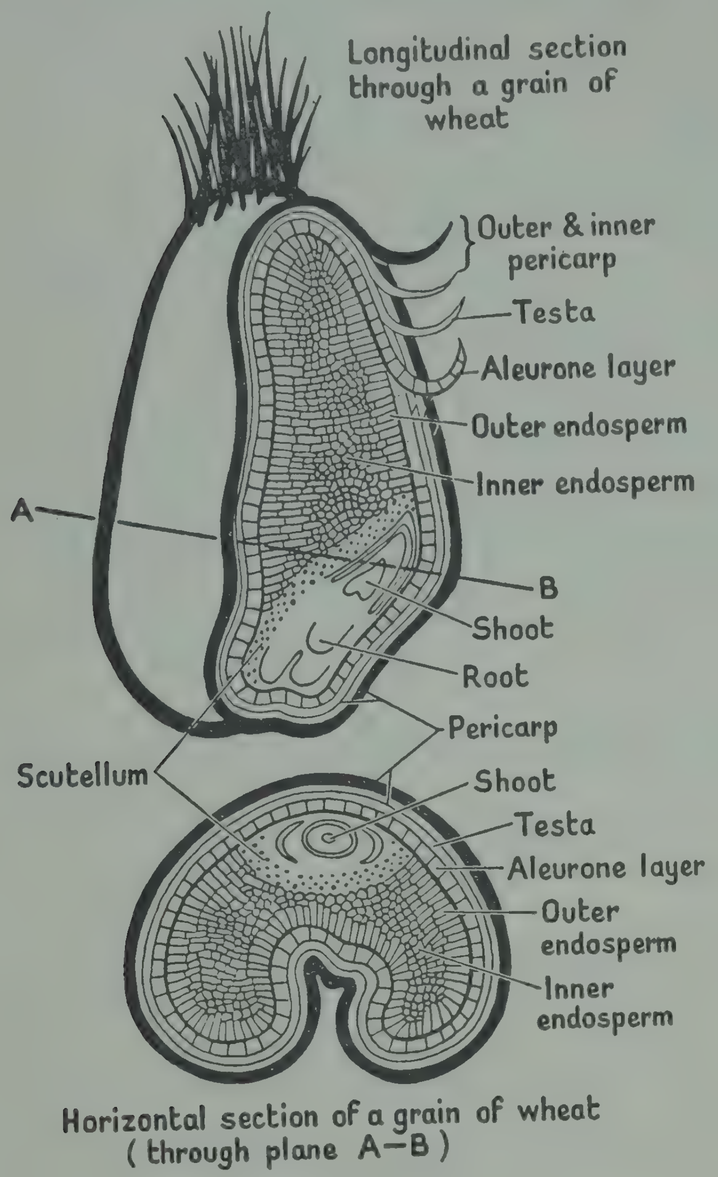

Longitudinal section through a grain of wheat
Outer & inner pericarp
Testa
Aleurone layer
Outer endosperm
Inner endosperm
A
B
Shoot
Root
Pericarp
Scutellum
Shoot
Testa
Aleurone layer
Outer endosperm
Inner endosperm
Horizontal section of a grain of wheat (through plane A—B)

FIG. 12, 7

R. A. McCance (1946), *Lancet*, i, 77.

In the process of milling the degree of extraction (the percentage of the whole grain retained in the flour) can be varied widely. 'Wholemeal' (92 per cent. extraction) or flour of 100 per cent. extraction contains a large proportion of indigestible fibrous matter. Roller-milling usually gives only 70 to 73 per cent. extraction, but the figure can be considerably increased, and since 1953 the National Flour of Great Britain has been of 80 per cent. extraction. Flour of this quality contains a high proportion of the minerals and vitamins of the original wheat grain but also much phytic acid which, as we have seen, may interfere with absorption. When the figure falls below 80 per cent. the loss in minerals and vitamins is considerable (see Table 12, 6).

The nutritive value of bread baked from white flour can also be artificially increased by 'fortification,' that is by the addition of vitamins and inorganic salts, and in Britain calcium carbonate is now added to National Flour at the rate of 14 oz. per 280 lb. sack.

TABLE 12, 8

Composition of Different Parts of the Wheat Grain

	Per cent. by weight (approx.)	Thiamine μg. per g.	Riboflavin μg. per g.	Nicotinic acid μg. per g.	Iron mg. per 100 g.
Whole wheat	100	3·6	1·6	50	3·5
Clear bran, including aleurone layer	12	4·8	5	250	12
Scutellum	1·5	165	15	60	9
Embryo	1	9	15	60	9
Outer endosperm	3	4·5	1·8	150	10
Bulk endosperm, including outer endosperm	85·5	0·6	0·7	22	2·1

From T. Moran & Sir Jack Drummond (1945), *Lancet*, ii, 698.

Bread.—When flour is kneaded with water the gliadin and the glutelin form the gluten of the dough which is puffed up by carbon dioxide produced by the action of yeast on carbohydrate. During baking the mass forms a sponge, the starch grains are burst by the heat, and some of the starch is converted into soluble starch and dextrins, and in the crust into brown caramelized products. The composition of bread is shown in Table 12, 6. In the making of cakes the carbon dioxide is produced by the action of tartaric acid on sodium bicarbonate. When a slice of bread is *toasted* much water is lost and the starch in the surface layers is largely converted into dextrins.

TABLE 12, 9

Composition of Fruits and Vegetables (edible portion)

	Per 100 g.						
	Water g.	Protein g.	Fat g.	Carbohydrate g.	kcal	Calcium mg.	Iron mg.
Apples, eating	85	0·3	0·0	8·6	36	4	0·3
Bananas	71	1·1	0·0	17·3	74	7	0·4
Black currants	77	0·9	0·0	5·9	27	60	1·3
Oranges	86	0·8	0·0	7·7	34	41	0·3
Tomatoes	93	0·9	0·0	2·5	14	13	0·4
Prunes	23	2·4	0·0	36·3	155	38	2·9
Raisins	22	1·1	0·0	58·0	236	61	1·6
Almonds	5	20·5	53·5	3·9	579	247	4·2
Beans, broad	77	7·2	0·5	9·5	71	30	1·1
Cabbage	92	1·5	0·0	5·0	26	65	1·0
Carrots	90	0·7	0·0	4·9	22	48	0·6
Lettuce	95	1·1	0·0	1·6	11	26	0·7
Parsley	79	5·2	0·0	0·0	21	325	8·0
Peas, green	79	5·8	0·0	9·5	61	15	1·9
Potatoes	78	2·0	0·0	16·2	73	8	0·7
Spinach	91	2·7	0·0	2·5	21	..	..
Soya bean, full fat flour	7	40·4	23·5	13·3	335	253	9·7

Oatmeal is a cereal particularly rich in protein, in fat and in minerals. A high proportion of its phosphorus, however, is not available because it is present as phytic acid which is not absorbed. In addition, phytic acid tends to combine with calcium and so retards the absorption of this important mineral. This disadvantage can be to some extent overcome by taking porridge with a liberal supply of milk, as is the usual custom.

Fruits and Vegetables.—Most green vegetables contain a large proportion of indigestible fibrous material, but they are important sources of minerals and of ascorbic acid and carotene, the precursor of vitamin A. Figures for some common vegetables are shown in Table 12, 9.

Potatoes contain large amounts of starch in the form of granules which swell up and burst on cooking. They also supply a large proportion of the total ascorbic acid intake of most people, not because they are especially rich in the vitamin but because they are eaten regularly in large amounts. They are also an important source of iron. Carrots are especially rich in the provitamin carotene. Turnips are of little nutritive value but are moderately rich in ascorbic acid. The pulses (beans, peas, lentils) are important sources of protein (20 to 25 per cent. in the dried state). In the fresh form (but not when dried) they supply ascorbic acid and carotene. The soya bean has a particularly high protein and fat content (Table 12, 9). Most fruits are of little nutritive value except as sources of ascorbic acid. When ripe they contain varying amounts of sugar and the banana is unique in supplying starch as well. The tomato is a good source of ascorbic acid and carotene. Dried fruits contain little ascorbic acid but their sugar provides calories. Nuts are rich in fat and in protein and are good sources of the vitamin B group.

TABLE 12, 10

Composition of Miscellaneous Foodstuffs

	Per 100 g.						
	Water g.	Protein g.	Fat g.	Carbo-hydrate g.	kcal	Calcium mg.	Iron mg.
Honey	23	0·4	0·0	68·8	277	5	0·4
Chocolate, milk	1	7·4	34·1	47·1	525	175	1·7
Chocolate, plain	0	4·6	32·5	52·8	522	26	3·3
Cocoa	3	20·4	25·6	31.5	438	51	14·3
Jams	29	0·2	0·0	62·1	249	12	..
Beer	..	0·2	0·0	1·7	29–38	10	0·1

Foods such as chocolate and cocoa (Table 12, 10) have a relatively high protein, fat and carbohydrate content and have thus a high energy value, but beverages such as tea and coffee are of little calorific value unless taken with milk and sugar. Tea and coffee contain the cerebral stimulant caffeine. Cocoa is a valuable source of iron.

Although *alcohol* has a calorific value of 7 kcal per g. its value as a food is restricted by its pharmacological action and by the fact that the body can metabolize only about 10 ml. per hour. Beer contains, in addition to about 4 per cent. of alcohol, small amounts of carbohydrate and significant amounts of calcium and of members of the vitamin B group derived chiefly from the yeast used in its fermentation.

Most foods of plant origin contain larger or smaller amounts of indigestible matter, chiefly cellulose, which is usually known as *roughage*. Although small quantities may be digested by bacterial action in the large intestine of man, the chief value of cellulose in the diet is to increase the bulk of the intestinal contents and so stimulate peristalsis (p. 282).

EFFECTS OF COOKING

The chemical composition of food, its digestibility and its palatability may be considerably altered in the process of cooking which consists essentially in the application of heat either directly as in grilling, baking and roasting, or indirectly by means of hot water (boiling and steaming) or of hot fat (frying). Such treatment, apart from rendering most foodstuffs more palatable and more digestible, has the effect of destroying bacteria.

The most obvious effect of cooking on protein is coagulation, such as occurs during the boiling of an egg. When meat is cooked water is lost, soluble protein is coagulated and collagen is converted to gelatin. At the same time the fibres shrink and become softer and looser so that the meat is easier to chew. The digestibility of protein is thus improved by moderate cooking but it is decreased by overheating.

The starch granules of cereals and potatoes swell up and burst during cooking and the starch is obtained in a form which is easily digested and absorbed. In the cooking of vegetables the cellulose framework is softened and loosened. During the process of boiling, soluble materials such as vitamins and salts are dissolved out of the food and may be lost unless the water is used to make soup, or employed in such small quantity that little extraction occurs. Loss of soluble materials is much less during roasting and baking, and less still during grilling and frying, in which the heat is usually applied for a much shorter time. The mineral content of food is affected little by cooking, but nutritionally important amounts of calcium may be added to the diet when hard water is used in cooking or for drinking, and significant amounts of iron may be dissolved from cooking utensils.

Cooking has little or no effect on any of the vitamins except thiamine and ascorbic acid. Thiamine is destroyed by heating, especially in the presence of alkali, and does not survive the temperature employed for example in the baking of biscuits. It is, however, not destroyed in the baking of bread unless baking soda is used. As much as 60 per cent. of the ascorbic acid content of foods may be lost during cooking because (*a*) it is destroyed by the enzyme ascorbic acid oxidase (p. 138) during the preparation of vegetables, (*b*) it is very soluble and thus is easily dissolved out by the water in which vegetables are cooked, and (*c*) part of it is destroyed by heat, especially in an alkaline medium. Vegetables should therefore be prepared immediately before they are to be used and cooked for as short a time as possible in the minimum of water.

Waste.—It must not be forgotten that there may be a considerable difference between food as purchased by the housewife and the food actually eaten by her family, and a due allowance must always be made for waste. Certain portions of food as purchased are normally discarded as inedible, for example bones, bacon-rind, the outer leaves of cabbages and other leafy vegetables and potato-peelings. These parts amount to a large proportion of the food bought, and to allow for this the figures quoted in Tables 12, 4, to 12, 9, refer to the edible portions of the foods only. In addition there is often a certain amount of edible residue left in cooking utensils and on plates after a meal. In assessing the value of any diet, therefore, it is important to measure as nearly as possible the food eaten by the individuals concerned rather than the food supplied to the kitchen. If this cannot be done, an allowance of 10 per cent. may be made for cooking and plate wastage. This is probably a liberal estimate and, in any case, the figure varies widely according to circumstances.

There can be no doubt that the stimulation of the appetite, the selection, cooking and attractive serving of food, together with the circumstances in which the meals are taken, the feelings of satisfaction following a meal and the idiosyncrasies of the eater are all factors of importance in assessing the nutritional value of a diet.

THE 'NORMAL' DIET

We have already seen that all our foodstuffs consist of some or all of the six chief nutrients: carbohydrates, fats, proteins, water, minerals and vitamins.

The carbohydrates and fats and such protein as is not used for growth and repair provide calories, the requirements of which in different circumstances have been dealt with in Chapter 11. The vitamins, the minerals and water have likewise been considered elsewhere. In the following section we shall discuss how much protein, fat and carbohydrate are required in the diet of a normal human being in ordinary circumstances.

Protein Requirements.—Protein is one of the most important nutrients. In the growing child an abundant supply of protein is required in the diet to provide the building stones (amino-acids) out of which new tissue protein is constructed in the process of growth; in the diet of the adult protein is required to maintain the proper level of body protein, to make good tissue lost by wear and tear, to build up new tissue protein after a wasting illness, and to supply amino-acids essential for the synthesis of enzymes and certain of the hormones, for example the tyrosine required for the synthesis of adrenaline and thyroxine.

We have already seen that the protein molecule consists of long chains of amino-acids joined together by peptide linkages. In the process of digestion these chains are broken down and the constituent amino-acids are absorbed from the gut and carried to the liver, where each is separately dealt with (see p. 339). According to the type of protein eaten, the proportions of the various amino-acids reaching the tissues vary within wide limits, and the nutritional value of any particular protein depends to a very large extent upon the nature of its constituent amino-acids and their relative proportions. Some of the amino-acids can be readily synthesized by the body from ammonia and simple carbon compounds; their presence in the diet is therefore not essential. Other amino-acids required by the body cannot be synthesized by the tissues and must be supplied in the diet. These are known as the *essential* or *indispensable amino-acids*. In some cases, for example lysine, the whole molecule must be supplied in the diet ready made, but in others a particular grouping, or nucleus, is all that is necessary which the body can build upon to form the appropriate amino-acid. For example, histidine is an essential amino-acid but the body can complete its synthesis if the iminazole ring is supplied in the form of iminazole pyruvic acid, the keto acid corresponding to histidine. This matter is discussed further on p. 351.

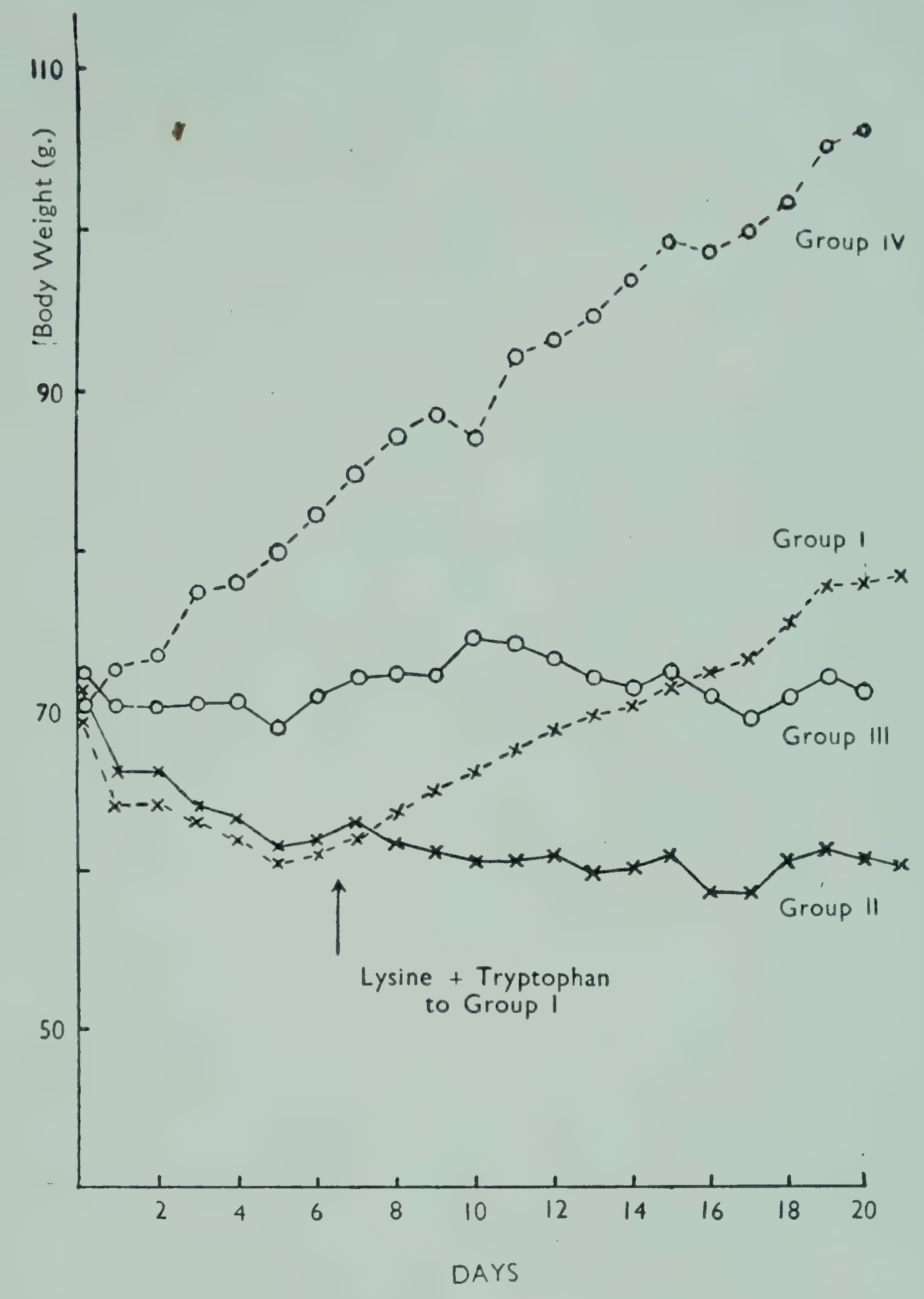


FIG. 12, 11

An experiment carried out by H. N. Munro to show the effect of adding lysine and tryptophan to a zein diet fed to young rats. The rats in Group II were fed on a diet containing 15 per cent. zein as sole source of protein. They did not grow and tended to lose weight. The rats in Group I were similarly fed but lysine and tryptophan were added at the point indicated by the arrow. Good growth then occurred. The rats in Group III were pair-fed with those in Group II on a diet containing 15 per cent. casein in place of zein, that is they were given only that amount of food which the rats in Group II ate spontaneously. They showed poor growth, but they did not lose weight. The rats in Group IV received the casein diet in the same amounts as those in Group III but they were allowed to eat a non-protein diet *ad libitum* as well. They showed excellent growth. It is clear that, while addition of the missing amino-acids improves the deficient diet, the animals on the deficient diet show poor appetite and this tends to aggravate their condition.

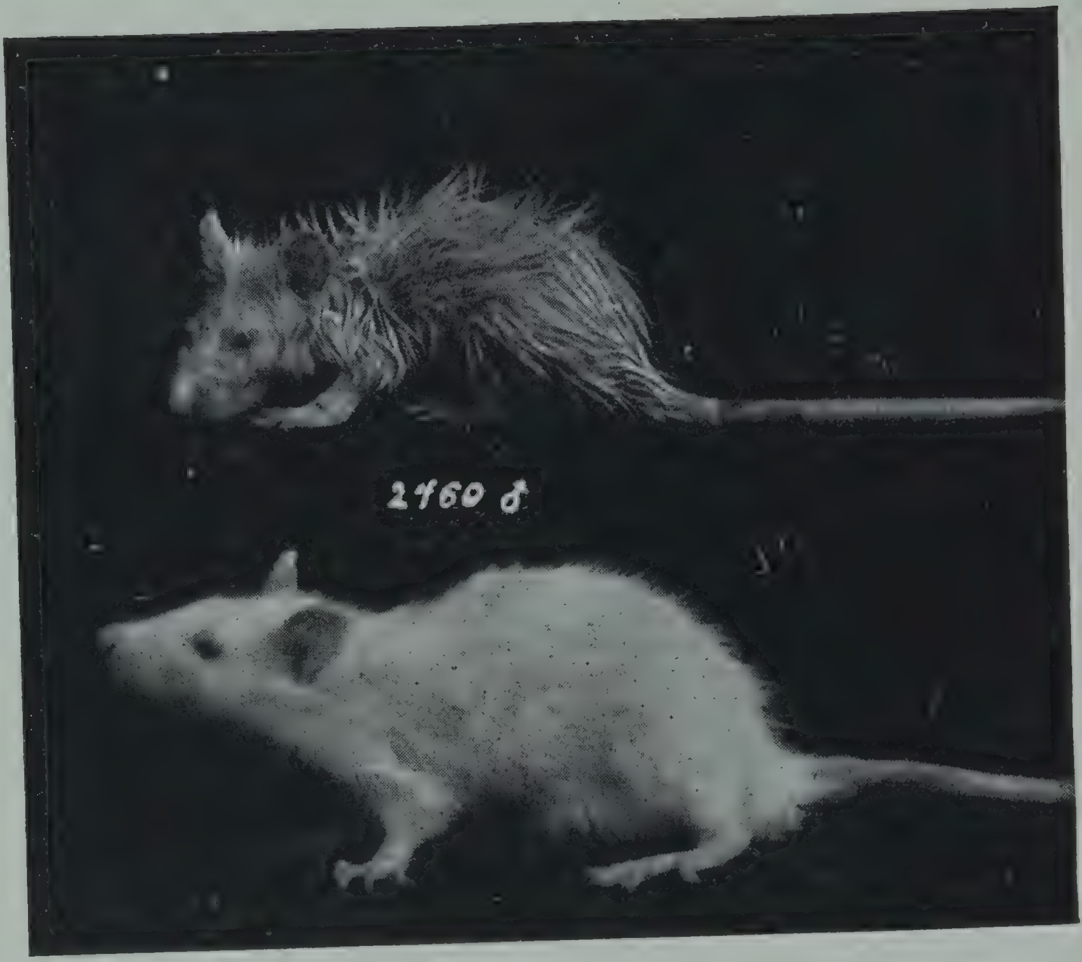


FIG. 12, 12

The effect of the absence of valine from the diet of the rat. The upper photograph shows a rat on the 28th day of valine deprivation, when the animal had lost one-third of the body weight, was too weak to eat, and appeared at the point of death. A solution of valine was administered by a pipette and when the animal had sufficiently recovered the valine was included in the diet. The lower photograph shows the same rat after valine had been administered for 25 days (W. C. Rose & A. Eppstein (1939), *J. biol. Chem.* **127**, 677).

It is important to understand that the division between the essential and the non-essential amino-acids is not rigid. Thus the amino-acid glycine is readily synthesized by the tissues but the glycine requirements of the growing chick are so great that the rate of utilization exceeds the rate of synthesis and glycine becomes one of the essential amino-acids in its diet. It should also be noted that if certain essential amino-acids are present in suboptimal amounts the deficiency in the diet may be made good by others not normally essential. Thus methionine is an essential amino-acid part of which may be converted to cystine in the body (p. 353) and used in reactions involving cystine. Although cystine cannot be transformed to methionine its presence in the diet releases dietary methionine for other purposes. To take another example, the tissues can synthesize tyrosine from phenylalanine but not phenylalanine from tyrosine. Phenylalanine is therefore an essential amino-acid while tyrosine is not, but the addition of tyrosine to a diet containing suboptimal amounts of phenylalanine may prevent the appearance of deficiency symptoms.

The amino-acid requirements of the young growing animal may differ in some respects from those of the adult. In the growing animal the importance of certain amino-acids may easily be demonstrated by observing the effects of their removal from the diet.

For example, in the well-known experiments of Osborne and Mendel young rats were fed on an otherwise adequate diet in which the only source of protein was zein (from maize), which is deficient both in tryptophan and in lysine. On this diet they not only ceased to grow but lost weight. When tryptophan was added the fall in weight was arrested; but when tryptophan and lysine were added normal growth was resumed. Tryptophan and lysine are obviously therefore essential amino-acids for the growing rat (Fig. 12, 11). The question of amino-acid requirements has been thoroughly examined by Rose who used mixtures of the known amino-acids in place of whole proteins. By this means, amino-acids could be withdrawn one by one and the effect on growth and on nitrogen balance observed. Those interfering with growth or causing negative nitrogen balances on withdrawal were classified as essential amino-acids. Conversely, an amino-acid which could be removed without affecting growth was non-essential (dispensable). Rose has defined an essential amino-acid as one which "cannot be synthesized by the animal body from materials ordinarily available at a speed commensurate with the demands for normal growth". An amino-acid is regarded as essential if none can be synthesized or if only an insufficient amount can be synthesized in the body.

The effect of removing the amino-acid valine from the diet of the young rat is shown in Fig. 12, 12. In the absence of valine, growth ceases, the animal's weight falls rapidly and it becomes ataxic. It is now known that ten amino-acids are indispensable for the growth of the young rat. They are lysine, tryptophan, phenylalanine, leucine, isoleucine, threonine, methionine, valine, histidine and arginine. It is probable that they are also essential for the growth of the child, although the human requirements have not yet been accurately worked out.

The amino-acid requirements of the adult are not so easily determined, but it can be shown that, when certain of the amino-acids are missing from an otherwise adequate diet, the output of nitrogen in urine and fæces exceeds the intake, that is there is a negative nitrogen balance (p. 188) which can be abolished by the addition of the appropriate amino-acid to the diet. Such experiments are usually carried out with diets in which protein has been replaced by artificial amino-acid mixtures. Instead of nitrogen balance, body weight may be used as an index. Thus it has been shown that elimination of lysine from the diet of the adult rat is followed by a steady fall in body weight (Fig. 12, 13).

It has recently become possible to accumulate sufficient stocks of pure amino-acids to enable tests to be made on human subjects fed on diets containing mixtures of pure amino-acids in place of protein. It has been found that the amino-acids essential for the maintenance of nitrogen balance are lysine, tryptophan, phenylalanine, leucine, isoleucine, threonine, methionine and valine. The position of histidine and arginine is less clear. They are apparently not essential for nitrogen equilibrium in man but their absence may have specific effects. Thus the absence of arginine from the diet is followed by a reduction of the number of spermatozoa in the seminal plasma. The remaining nine amino-acids (alanine, aspartic acid, citrulline, cystine, glutamic acid, glycine, hydroxyproline, serine, tyrosine) are certainly not required by either man or the growing rat.

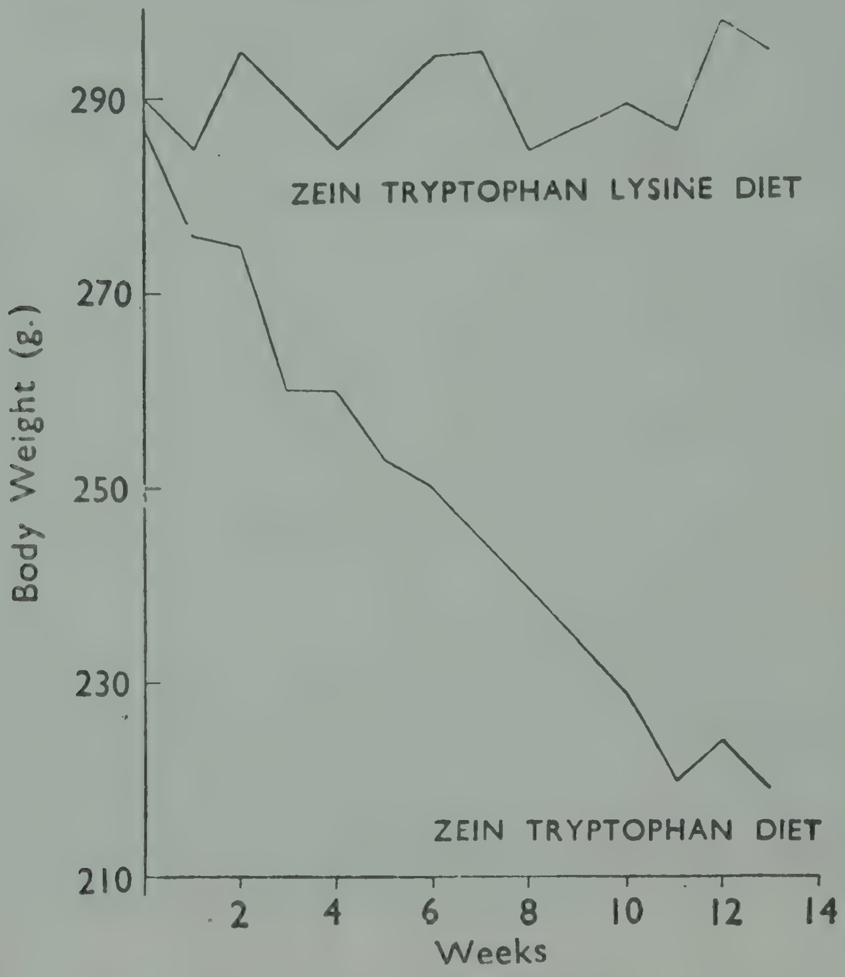


FIG. 12, 13

Weight changes of adult rat on zein diets supplemented with tryptophan with and without added lysine (A. Neuberger & T. A. Webster (1945), *Biochem. J.* **39**, 200).

It is now obvious that the dietary value of any protein depends upon its content of essential amino-acids. A protein such as lactalbumin containing all the essential amino-acids is clearly nutritionally superior to one, such as zein of maize, which is deficient in tryptophan and lysine, or gelatin, which is deficient in both tryptophan and tyrosine. In other words these proteins have different biological values. Proteins of high biological value are capable of supporting life and maintaining nitrogen balance (p. 188) if they are the sole source of nitrogen in the diet. They supply *all* the essential amino-acids and it is important to note that most animal proteins with the exception of gelatin fall into this category. Most plant proteins, on the other hand, are of low biological value. They cannot support life if they are the sole source of nitrogen in the diet, they are lacking in one or more of the essential amino-acids, and they are said to be less well digested and absorbed than the animal proteins. The distinction between these two types of protein is however largely artificial. Man does not normally eat a diet containing a simple protein, and in a mixed diet any amino-acid missing from one protein may be abundantly supplied by another. It is thus quite possible to maintain nitrogen equilibrium on a diet consisting of a mixture of vegetable proteins. However, it is usually recommended that at least 50 per cent., and preferably 60 per cent., of the protein in a mixed diet should be of animal origin.

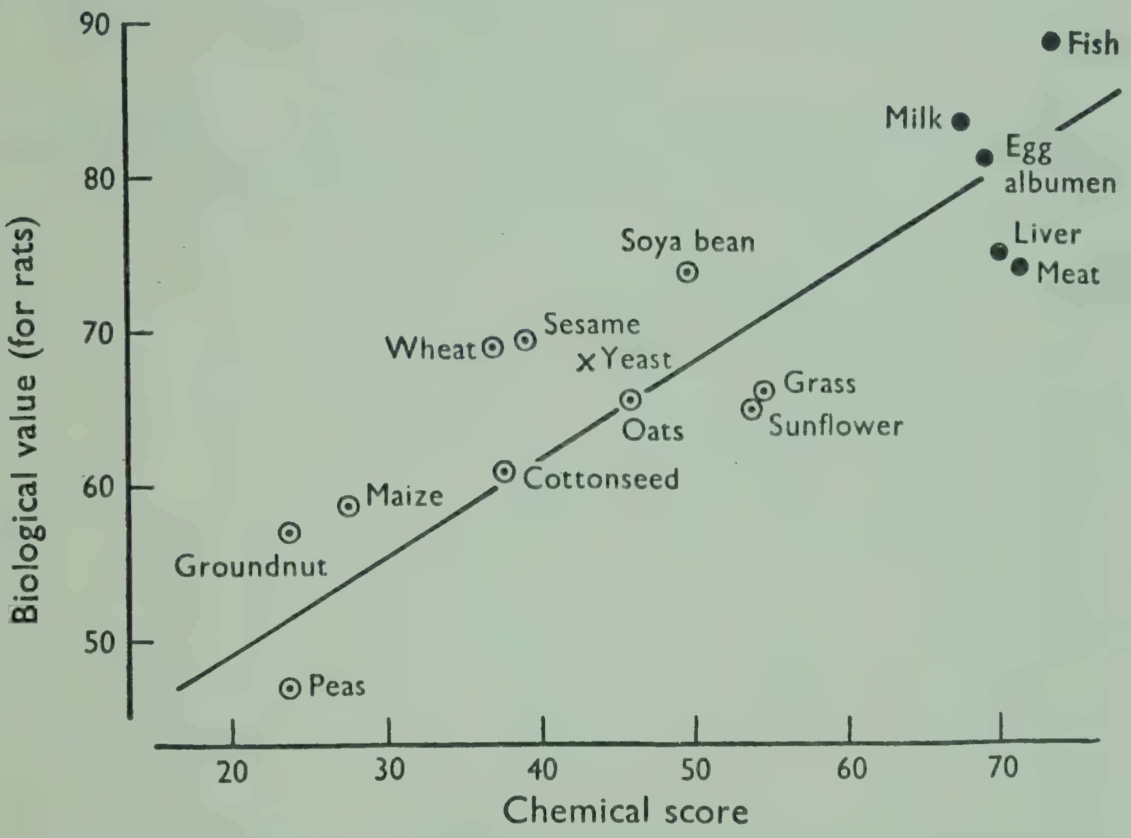


FIG. 12, 14

Correlation diagram of the biological value and chemical score of sixteen animal (●) and vegetable (⊙) materials. (From K. J. Carpenter (1951), *Brit. J. Nutr.* **5**, 243.)

A recent method of expressing the nutritional value of a protein is due to Block and Mitchell, who calculate a *chemical score* in terms of whole egg protein which is almost wholly utilized by the young rat. Fairly reliable figures are available for the essential amino-acid content of most common feeding stuffs and the chemical score is obtained by calculating the concentration of each essential amino-acid in the protein of a foodstuff as a percentage of the concentration of the same amino-acid in whole egg protein. The lowest percentage obtained is used as the chemical score since the limiting amino-acid is held to determine the value of the whole protein. Fig. 12, 14 shows the chemical score of various proteins plotted against their biological value for rats as determined by nitrogen balance methods to be described below. There is a high degree of correlation between the two methods of evaluation and it is clear that animal materials are superior to those of plant origin.

The minimum protein requirement can be assessed by nitrogen balance experiments. On any given diet, there are three possible results: (i) nitrogen output may exceed intake; the subject is then in negative balance; (ii) output may be less than intake, in which case he is storing nitrogenous compounds and is said to be in positive balance; (iii) nitrogen intake may be just adequate to offset excretion and the subject is then in nitrogen equilibrium. This technique is not limited to growing animals, and moreover it can be applied to human subjects over short periods. If a human subject or an experimental animal is given a diet containing no protein, the output of nitrogenous compounds in the urine and fæces is reduced not to zero, but to a low value (Fig, 12, 15). This is referred to as the 'endogenous output of nitrogen,' and means that continuous breakdown of tissue is going on. Since the intake of protein is zero, the subject is in negative nitrogen balance to the extent of the nitrogen lost in urine and fæces. A protein containing an amount of nitrogen equal to this negative balance is now added to the diet. If this protein contains amino-acids in the optimal proportions for utilization by the body, it will completely eliminate the negative nitrogen balance. This is found to apply to the mixed proteins of whole egg when fed to the rat. If, on the other hand, the protein is not of the same quality as whole egg protein, rather more will be needed to restore nitrogen equilibrium owing to lack of one or more of the essential amino-acids (Fig. 12, 15). Further, the addition of an incomplete protein like zein to the diet always leads to negative nitrogen balance, even when large amounts are added, because of the complete absence of tryptophan

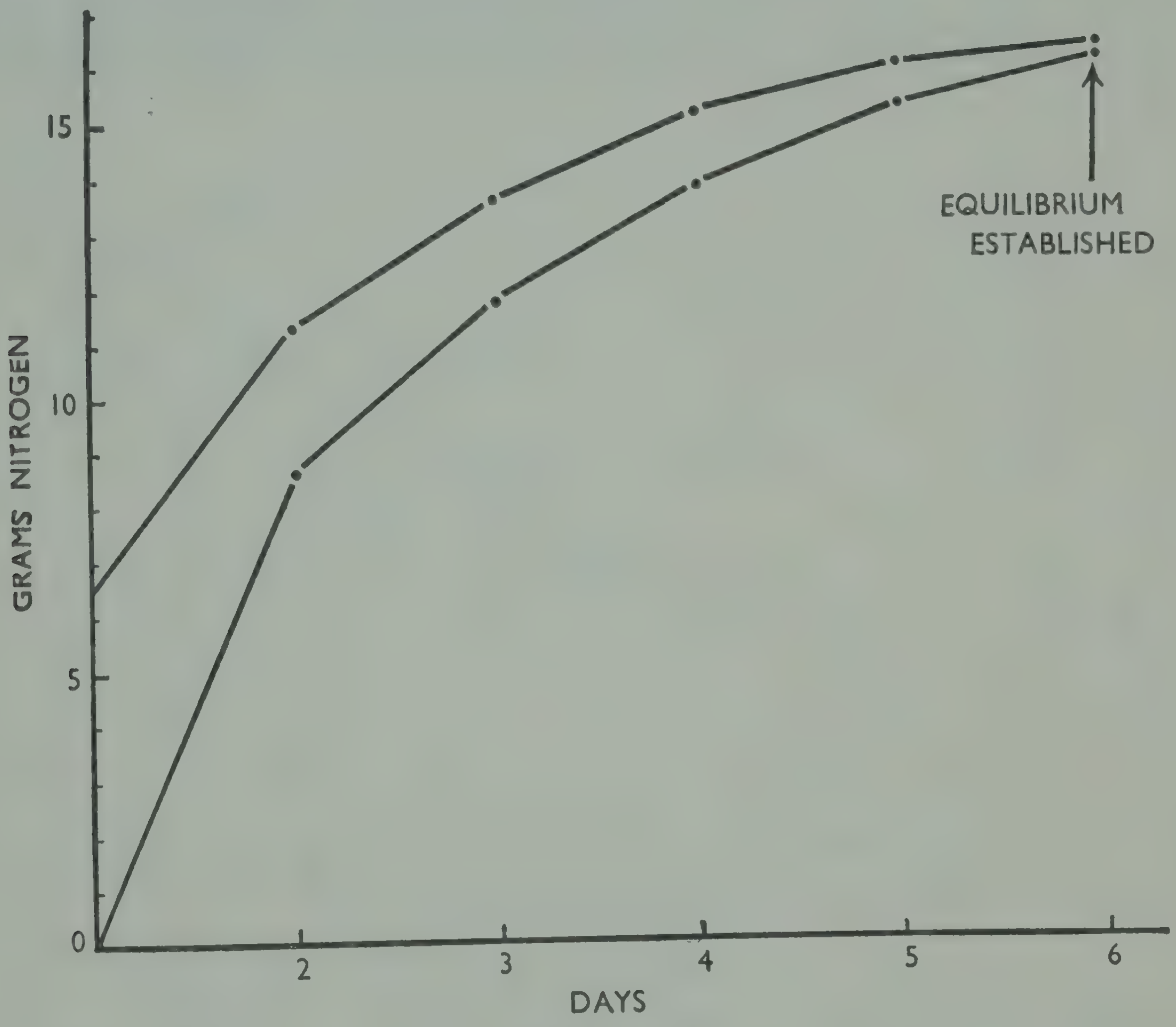


Fig. 12, 15

Establishment of nitrogen equilibrium in a dog fed increasing amounts of protein. When the protein intake is zero the nitrogen output is about 6·5 g. per day. As the protein intake is increased the nitrogen output increases also, but at a slower rate, until intake and output finally balance at about 16·5 g. nitrogen per day. Upper curve, output. Lower curve, intake.
(From Parsons *after* Voit.)

and lysine. The nitrogen balance procedure has been used by Mitchell to determine the biological value of proteins by making use of the formula:

$$\text{Biological value} = \frac{\text{Dietary N retained}}{\text{Dietary N absorbed}} \times 100$$

Thus in the case of whole egg protein, fed to rats at a level of 30 mg. absorbed protein nitrogen per day, nitrogen balance improved by 30 mg. per day and thus all the absorbed nitrogen was retained, giving a biological value of 100. On the other hand, beef protein fed at a level of 132 mg. nitrogen daily and wholly absorbed, gave rise to an improvement in nitrogen balance of 102 mg. nitrogen, that is, a biological value of 77. Some results obtained by this method are shown in Table 12, 16.

TABLE 12, 16

The Biological Value of some Proteins for Growing Rats, Mature Rats and Adult Man (after H. H. Mitchell)

Protein	Biological value		
	Growing rat	Mature rat	Adult man
Egg albumin	97	94	91
Beef muscle	76	69	67
Wheat gluten	40	65	42
Caseinogen	69	51	56

If the diet contains carbohydrate and fat for the provision of energy much less protein is required to produce nitrogen equilibrium than when protein alone is fed. Carbohydrate and fat are therefore said to exert a *protein-sparing action*, and the smallest amount of protein needed in the diet for the attainment of nitrogen equilibrium when energy requirements are being adequately met by carbohydrate and fat gives us an estimate of the minimum protein requirement. It may be surprisingly low; in a classical series of experiments on himself Chittenden was able to maintain nitrogen balance on a diet containing much less protein than is usually consumed by the adult man. It is likely however that there is a considerable difference between the minimum protein requirement and the optimum requirement. If his calorie demands be met, man can adjust himself to wide variations in the level of protein intake. In temperate climates he usually elects to take between 10 and 12 per cent. of his total calorie intake in the form of protein, of which about 60 per cent. is animal protein. Some races take more protein than this without any apparent deleterious effect. The Eskimo, for example, may take as much as 40 per cent. of his total calorie intake in the form of protein. One factor tending to limit the intake of protein is its price since it is a relatively expensive article of diet. In Britain the cheapest sources of protein are milk, cheese and herring.

TABLE 12, 17

Summary of Daily Dietary Allowances as Recommended by the Committee on Nutrition of the British Medical Association (1950)

Age and Sex	Description	kcal per day	Protein	Iron	Calcium	Vitamin A and Carotene	Vitamin D	Thiamine	Nicotinic acid	Riboflavin	Vitamin C
			g.	mg.	g.	i.u.	i.u.	mg.	mg.	mg.	mg.
Both—											
0–1	..	1000	37	6·5	1·0	3000	800	0·4	4	0·6	10
2–6	..	1500	56	7·5	1·0	3000	400	0·6	6	0·9	15
7–10	..	2000	74	10·5	1·0	3000	400	0·8	8	1·2	20
Males—											
11–14	..	2750	102	13·5	1·3	3000	400	1·1	11	1·6	30
15–19	..	3500	130	15·0	1·4	5000	400	1·4	14	2·1	30
20 +	No work, almost basal (e.g. lying in bed)	1750	51	12·0	0·8	5000	..	0·7	7	1·0	20
	Sedentary work (30 kcal/hr) and little travelling (65 kcal)	2250	66	12·0	0·8	5000	..	0·9	9	1·4	20
	Light work (70 kcal/hr) and travelling (130 kcal)	2750	80	12·0	0·8	5000	..	1·1	11	1·6	20
	Medium work (100 kcal/hr) and travelling (130 kcal)	3000	87	12·0	0·8	5000	..	1·2	12	1·8	20
	Heavy work (200 kcal/hr) and travelling (130 kcal)	3500	102	12·0	0·8	5000	..	1·4	14	2·1	20
	Very heavy work (300 kcal/hr) and travelling (130 kcal)	4250	124	12·0	0·8	5000	..	1·7	17	2·6	20
	Extremely heavy work (450 kcal/hr) and travelling (130 kcal)	5000	146	12·0	0·8	5000	..	2·0	20	3·0	20

Females—											
11–14	..	2750	102	13·5	1·2	3000	400	1·1	11	1·6	30
15–19	..	2500	93	15·0	1·1	5000	400	1·0	10	1·5	30
20 +	No work, almost basal (e.g. lying in bed)	1500	44	12·0	0·8	5000	..	0·6	6	0·9	20
	Sedentary work (30 kcal/hr) and little travelling (50 kcal)	2000	58	12·0	0·8	5000	..	0·8	8	1·2	20
	Light work (70 kcal/hr) and travelling (100 kcal)	2250	66	12·0	0·8	5000	..	0·9	9	1·4	20
	Medium work (100 kcal/hr) and travelling (100 kcal)	2500	73	12·0	0·8	5000	..	1·0	10	1·5	20
	Heavy work (200 kcal/hr) and travelling (100 kcal)	3000	87	12·0	0·8	5000	..	1·2	12	1·8	20
	Very heavy work (300 kcal/hr) and travelling (100 kcal)	3750	109	12·0	0·8	5000	..	1·5	15	2·2	20
Pregnancy—											
First half	..	2500	93	12·0	0·8	6000	400	1·0	10	1·5	40
Second half	..	2750	102	15·0	1·5	6000	600	1·1	11	1·6	40
Lactation	..	3000	111	15·0	2·0	8000	800	1·4	14	2·1	50

The British Medical Association (1950) recommendations are in general similar to those of the National Research Council of the U.S.A. (1948) but differ slightly in the recommendations for protein.

The assessment of the optimum level of protein intake can be made only on arbitrary grounds at present and several standards have been proposed. The Health Committee of the League of Nations in 1936 proposed that the protein intake of the adult human subject should be about 1 g. per kg. body weight per day. For a man weighing 70 kg. the daily intake would thus be 70 g. protein, and this is the figure proposed in the recommendation of the National Research Council in Washington in 1948. This report suggests 70 g. protein per day for men, 60 g. for women with an increase to 85 g. during pregnancy to meet the needs of the growing fœtus, and to 100 g. during lactation to compensate for the loss of protein in the breast milk. The growing child requires relatively more protein than does the adult, that is 3 to 4 g. per kg. body weight for infants. The child of 10–12 years requires as much as the adult man, while older children require even more. These requirements are based on American estimates. The Committee on Nutrition of the British Medical Association recommended in 1950 that for adults not engaged in hard physical work not less than 11 per cent. of the energy allowance should come from protein. For pregnant and nursing women, infants, children and adolescents 14 per cent. of the caloric intake should be provided by protein (Table 12, 17).

In connexion with the protein requirements of the human subject several points must be kept in mind. The specific dynamic action (p. 169) of protein is high and the consumption of protein greatly increases the heat production of the body. This fact makes protein particularly suitable for people exposed to cold, and the inhabitants of Arctic regions, as we have seen, usually have a particularly high protein intake. Conversely, people who are taking a low protein diet tend to feel the cold. The high specific dynamic action of protein makes it desirable that the consumption of protein should be accompanied by the simultaneous consumption of carbohydrate and fat so that the energy intake be raised to allow for the increased metabolism caused by the protein. There is thus a good reason underlying the general habit of eating potatoes along with meat.

The popular impression that men engaged in hard physical work require more protein and consequently a higher meat ration than those engaged in sedentary occupations has very little scientific foundation. The fuel used in muscular work is carbohydrate, not protein, and the increased calorie needs of the manual worker can generally be met out of an increased consumption of carbohydrate and fat. But the man who is habitually engaged in hard physical work is generally larger, better developed and heavier than his sedentary comrade, and to that extent his protein requirements will be greater. It is of interest that the National Research Council recommends the same protein intake irrespective of physical activity. The British Medical Association Committee on Nutrition on the other hand relates the protein intake to the energy intake, so that a man engaged in very hard work should receive 150 g. protein (Table 12, 17).

Manifestations of Inadequate Protein Intake.—Inadequate intake of protein implies defective synthesis of body protein because important amino-acids may not be available in sufficient amounts. Hence its clinical manifestations may resemble those of specific deficiencies of individual amino-acids. Similar disturbances can be conditioned by excessive protein metabolism, as for example during illness or after injuries. When an animal is given a diet low in protein, the rate of loss of protein from different organs occurs with different speeds. The liver loses protein rapidly and extensively so that in a few days it may have lost 25 per cent. of its substance; with this goes evidence of impairment in liver function. Prolonged protein deficiency can cause the death of large numbers of the liver cells (liver necrosis). The kidney is less rapidly depleted of protein than the liver, and the loss of protein from skeletal muscle is much slower and less severe. Some organs, for example the eye, may actually continue to grow during protein depletion. Furthermore, a skin wound heals quite rapidly in a protein-deficient animal, though the process is less rapid than in the case of a well-fed animal. Malignant tumours also continue to grow in animals given protein-free diets. Since enzymes are protein in nature, the process of protein depletion causes a reduction in their concentration. The effect varies from enzyme to enzyme; thus liver phosphatase is hardly altered, whereas xanthine oxidase practically vanishes from the liver. Changes in plasma protein levels are important because, unlike the other manifestations of protein deficiency which have so far been mentioned, they can be measured in man. The effect of protein insufficiency is to cause a considerable reduction in the plasma proteins, notably in the albumin level. This hypoproteinæmia brings about a fall in the colloid osmotic pressure of the plasma and in consequence the subject may develop œdema by water escaping from the plasma into the tissues. A drop in the plasma volume thus occurs, masking the fall in the total plasma protein. It may, therefore, be more important to know the total circulating plasma proteins rather than their concentration; a low level of plasma protein in circulation, accompanied by a low plasma volume, gives a normal concentration.

Not only is the amount of circulating protein reduced in protein deficiency but also the capacity to respond to a call for more plasma protein. Whipple has shown that, after removing plasma (plasmapheresis) from a fasting dog previously fed on an adequate diet, the plasma protein level is restored within a short time (for example 24 hr). Since no food is given during this period, the building materials for the new plasma protein must come from the tissues. In protein deficiency, however, this response is much impaired. This can be demonstrated clinically by giving an intravenous injection of saline which increases the blood volume temporarily. If the subject is healthy the body responds by a rapid provision of plasma protein, so that the concentration does not fall as a result of the increased blood volume. If, on the other hand, the subject is depleted of protein, no such response occurs and the saline merely dilutes the plasma proteins with a resultant fall in concentration. The antibodies of the plasma which are defences against disease are protein in nature and are included in the globulin fraction. Although claims have been made that the ability to produce antibodies is impaired in protein deficiency, clinical evidence that the diet plays an important part in resistance to infection is meagre.

A particularly severe form of deficiency disease in undernourished children called *kwashiorkor* occurs in the tropics, notably in Africa where the protein of the diet is exclusively vegetable in nature. It affects children receiving essentially vegetable diets and is characterized by failure of growth, anæmia, hypoproteinæmia, œdema, dry scaly depigmented skin and later by severe liver damage. Kwashiorkor is the combined result of a number of deficiencies in which lack of amino-acids is probably the most important element. Treatment consists of modifying the diet to contain animal protein foods, especially milk.

It should be noted that immediately following physical injuries, such for example as a fractured bone, and after surgical operations, patients excrete more nitrogen than they take in, that is, they are in negative nitrogen balance. Cuthbertson has named this phenomenon the *catabolic response to injury*. Such patients therefore should receive a high-protein, high calorie diet to make good this loss.

Fat Requirements.—The fat of the diet is important not only on account of its high energy value but because it is the vehicle for the fat-soluble vitamins A, D, E and K, and contains the essential fatty acids *linoleic acid* and *linolenic acid*. Both are unsaturated acids which cannot be synthesized by the tissues. If they are absent from the diet of the rat symptoms of severe deficiency, characterized by cessation of growth and the development of a dry scurfy skin and scaly tail, make their appearance. The importance of these fatty acids in the human diet is however not fully established.

Provided that the supply of fat-soluble vitamins and essential fatty acids is adequate, fat itself is not an absolutely essential constituent of the diet. The fat intake in human dietaries varies over a wide range according to climate, race and dietary custom. In this country the fat intake is about 80 to 150 g. per head per day, and in Europe generally, and in America, the fat content of the diet usually accounts for 30 to 35 per cent. of the total calories. In tropical countries and in the East much smaller quantities of fat are normally eaten, while in circumpolar regions the fat intake may be very high indeed. The Eskimo can tolerate a daily intake of 300 g. of fat which would nauseate the average European. At the other end of the scale a diet lacking in fat tends to be bulky and unpalatable, since fat is used so extensively for baking, cooking, frying and for buttering bread. Although hard-and-fast rules cannot easily be established it can be said that when the intake is less than 3000 kcal per day fat should account for at least 25 per cent. of the total calories; if the intake is greater than 3000 kcal per day fat should account for at least 30 per cent. of the total calories. These figures give a minimum fat intake of 56 to 140 g. per day according to total requirements.

Carbohydrate Requirements.—When the requirements for protein and fat have been met, the remaining calories needed are derived from carbohydrate. The proportion of carbohydrate in the diet is subject to wide variations but in Europe and America roughly 56 per cent. of the total caloric intake comes from carbohydrate and it is usually recommended that not more than 66 per cent. should be obtained from this source. Carbohydrate foods are in general the least expensive articles of diet; hence carbohydrate tends to form a large proportion of the diet of the poorer sections of the community. In India and in China carbohydrate may account for as much as 78 per cent. of the total caloric intake.

It is of course desirable that sufficient carbohydrate should be present in the diet to prevent ketosis (p. 332), but the amount required for this purpose is very low and is in practice always exceeded unless deliberate measures are taken to supply a ketogenic diet.

A summary of the nutritional needs of the adult is given in the table of recommendations of the Committee on Nutrition of the British Medical Association (Table 12, 17).

The Effects of Starvation.—When the organism is deprived of food it calls first on its stores of fat and carbohydrate for the provision of energy. Stores of carbohydrate—glycogen in liver and muscle (p. 291)—are never great and are soon exhausted, but, in the well-nourished subject, fat stores may be sufficient to supply the energy needs of the body for several weeks. When fat is being metabolized in large amount, ketosis (p. 330) results and the urine may contain large amounts of ketone bodies and may have a high titratable acidity and ammonia content. Even before the reserves of non-nitrogenous fuel are exhausted, the tissue proteins begin to be broken down to provide energy and when the fat stores are exhausted protein is the sole source of energy. Thus, while nitrogen excretion is low in the early stages of starvation, it rises sharply after the exhaustion of carbohydrate and fat stores, when the tissue proteins are being drawn upon. This rise in nitrogen excretion heralds imminent death and is sometimes referred to as the *premortal rise.*

During starvation therefore, the body loses weight but the various organs and tissues are affected to different degrees. Fat may almost disappear from the body. While the spleen, liver and muscles lose some 50 per cent. of their weight, the heart and brain remain relatively little affected, losing but a small percentage of their substance.

MEALS

Having considered the requirements of the various nutrients in the diet, we may examine the composition of one or two sample meals to find if these requirements are being met and to see how they may be translated into terms of actual meals.

Table 12, 18 shows the composition of two meals. Meal 1 is a light lunch which might be consumed during the course of a morning's work; meal 2

TABLE 12, 18

Analysis of Two Meals *

MEAL 1	Weight oz.	kcal	Protein g.	Fat g.	Calcium mg.	Iron mg.
Roll	3·5	252	8·8	1·1	109	1·8
Butter	0·3	65	0·0	7·3	0	0·0
Cheese	2·0	234	14·2	19·6	460	0·4
Tea (with milk and sugar)	10·0	20	1·0	1·0	30	0·0
Total		571	24·0	29·0	599	2·2

* From *Manual of Nutrition*, 1951, Ministry of Food. London: H.M.S.O.

TABLE 12, 18—*continued*

MEAL 2	Weight oz.	kcal	Protein g.	Fat g.	Calcium mg.	Iron mg.
Mutton	2·4	225	8·9	21·1	7	1·4
Cabbage	3·0	21	1·2	0·0	54	0·9
Potato	4·0	84	2·4	0·0	8	0·8
Apples	3·0	36	0·3	0·0	3	0·3
Custard	2·0	62	1·8	2·2	70	0·0
Total		428	14·6	23·3	142	3·4

is of the type provided, for example, by a factory canteen as a midday dinner. A comparison of the two meals shows some striking differences. Meal 1, which appears superficially to be much less substantial than meal 2, is in fact much richer in calories, in protein, in fat and in calcium. It is also much richer in vitamin A on account of the high vitamin A content of butter and cheese. The only feature in which meal 2 is superior is in ascorbic acid content, derived largely from the cabbage.

Table 12, 19 shows the figures obtained for a whole day's diet by adding

TABLE 12, 19

Daily Diet of Women doing Light Engineering at a Factory in Birmingham in 1943 ¶

Meal	kcal	Protein g.	Fat g.	Calcium mg.	Iron mg.	Vitamin A i.u.	Thiamine mg.	Riboflavin mg.	Nicotinic acid mg.	Ascorbic acid mg.
Breakfast,* 7 A.M.	710	21	26	240	4	630	0·51	0·30	2·9	5
Snack,† 10 A.M.	210	9	8	140	1	230	0·12	0·12	0·6	0
Dinner,‡ 1 P.M.	830	30	27	350	4	1100	0·87	1·90	5·2	30
Snack,§ 4 P.M.	240	5	9	50	1	50	0·06	0·05	0·3	0
Tea,‖ 7 P.M.	570	19	21	100	4	560	0·51	1·12	3·9	31
Total day's nutrients	2560	84	91	880	14	2570	2·07	3·49	12·9	66

* Substantial breakfast. † Piece of cake. ‡ Meat, vegetables and a substantial pudding. § Scone. ‖ A good meal, including chipped potatoes.

¶ From *Manual of Nutrition*, 1951, Ministry of Food. London: H.M.S.O.

together the figures for individual meals of a group of women engaged in factory work. The total intake of 2560 kcal compares favourably with the allowance of 2500 kcal proposed in Table 12, 17 for women engaged in medium work. For such women the proposed protein allowance is 60 g. and the amount provided (84 g.) is therefore liberal. It accounts for 13 per cent. of the total energy. Fat provides 32 per cent. of the total energy, which is well over the minimum figure recommended on p. 194. Iron and calcium are both above the recommended allowances. The amount of vitamin A (2570 i.u.) is less than the recommended allowance of 5000 i.u. but is slightly greater than the minimum safe figure of 2500 i.u. Thiamine (2·07 mg.) and riboflavin (3·49 mg.) are both present in amounts exceeding the recommended allowances of 1·0 and 1·5 mg. respectively. Nicotinic acid (12·9 mg.) is above the allowance of 10 mg. and the amount of ascorbic acid (66 mg.) is ample. This diet is, on the whole, therefore, a satisfactory one.

REFERENCES

Valuable reviews of various problems of nutritional importance are to be found in the *Proceedings of the Nutrition Society* and in the *British Medical Bulletin*, 1944, Vol. 2, Nos. 10–11. The following publications will also be found useful:

ALBANESE, A. A. (1950). *Protein and Amino-Acid Requirements of Mammals.* New York: Academic Press.

BOURNE, G. H. & KIDDER, G. W. (1953). *The Physiology and Biochemistry of Nutrition.* New York: Academic Press.

CRUICKSHANK, E. W. H. (1952). *Food and Nutrition,* 2nd ed. Edinburgh: Livingstone.

DRUMMOND, J. C., & WILBRAHAM, A. (1939). *The Englishman's Food.* London: Cape.

FENELON, K. G. (1952). *Britain's Food Supplies.* London: Methuen.

Food Composition Tables for International Use (1949). Washington: Food and Agricultural Organization of the United Nations.

HOLDER, LORD, DODDS, E. C. & MORAN, T. (1954). *Bread.* London: Constable.

KEYS, A., BROZEK, J., HENSCHEL, A., MICHELSON, O. & TAYLOR, H. L. (1950). *The Biology of Human Starvation.* Oxford: University Press.

KITCHIN, A. H. & PASSMORE, R. (1949). *The Scotsman's Food.* Edinburgh: Livingstone.

MCCANCE, R. A. & WIDDOWSON, E. M. (1945). *Chemical Composition of Foods. Spec. Rep. Ser. med. Res. Coun., Lond., No.* 235. London: H.M.S.O.

Manual of Nutrition (1955, 4th ed.). Ministry of Food. London: H.M.S.O.

MORRISON, S. D. (1952). *Human Milk—Yield, Proximate Principles and Inorganic Constituents.* Farnham Royal: Commonwealth Agricultural Bureaux.

PYKE, M. (1952). *Townsman's Food.* London: Turnstile Press.

SAHYUN, M. (1948). *Proteins and Amino-Acids in Nutrition.* New York: Reinhold.

CHAPTER 13

TEMPERATURE REGULATION

WHEN calculated over a period of twenty-four hours the gross efficiency of the body (see p. 164) is relatively low. At least 95 per cent. of the 3000 kilocalories or so produced in the day appears as heat, the remainder appearing as mechanical energy. This heat serves to maintain the temperature of the body but it must be got rid of as fast as it is generated otherwise the body temperature will rise.

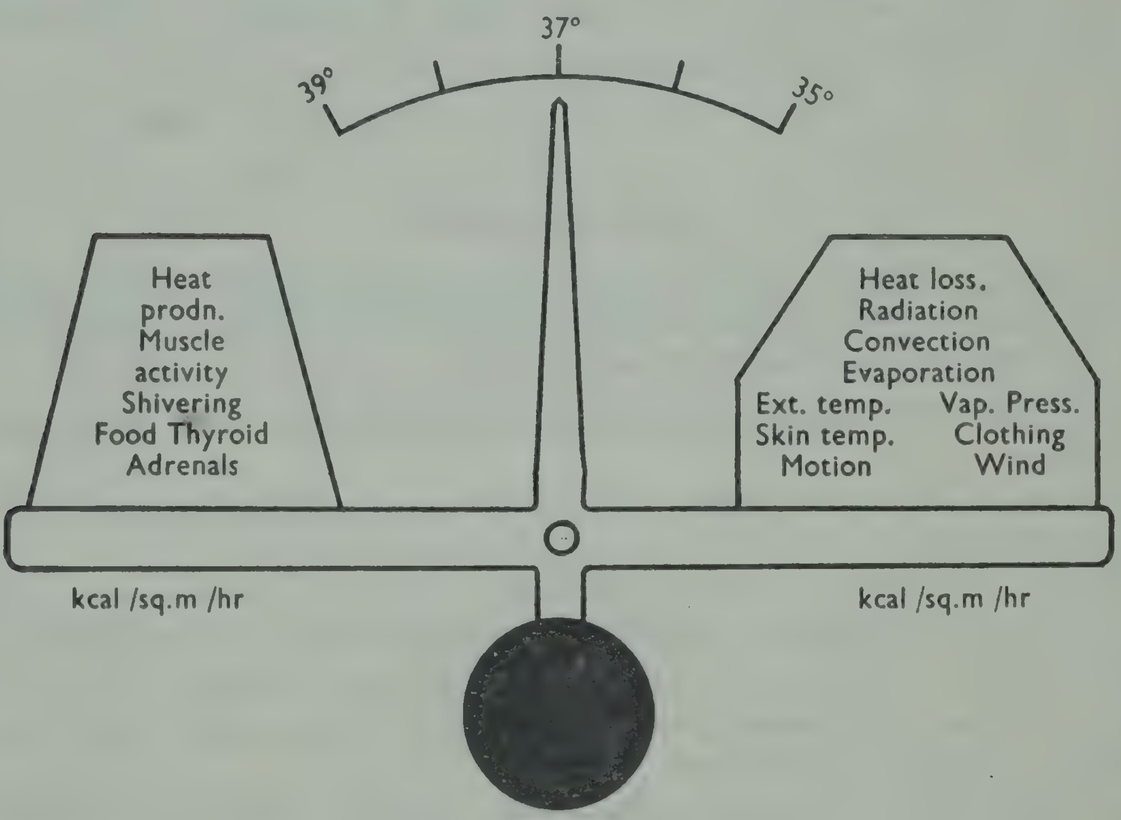

37°
39°
35°
Heat prodn. Muscle activity Shivering Food Thyroid Adrenals
Heat loss. Radiation Convection Evaporation
Ext. temp. Skin temp. Motion
Vap. Press. Clothing Wind
kcal /sq.m /hr
kcal /sq.m /hr

FIG. 13, 1

Diagram to illustrate the balance between heat production and heat loss. The black circle represents the heat capacity of the body.

(Modified from Du Bois.)

While the body temperature is remarkably constant in man, it is not so accurately regulated in other mammals. Without a precisely controlled temperature, accurate cerebration cannot occur; if the temperature-regulating mechanism is grossly upset cerebral functions are among the first to be seriously disturbed.

Many of the activities of the body are concerned with maintaining what Claude Bernard described as the fixity of the internal environment. Cannon called this the principle of homoiostasis, and temperature regulation is an important example of it. Animals which are able to maintain a relatively constant body temperature in spite of great variations of external temperature are said to be *homoiothermic*, whereas animals in which the body temperature varies with that of the environment are called *poikilothermic*, or cold-blooded. The former can obviously lead a more free and independent existence.

The body temperature is regulated by balancing heat loss against heat production. This balance (Fig. 13, 1), unlike those used in the laboratory, is insensitive to small changes. Because of the high thermal capacity of the tissues, sudden changes in heat production or heat loss are followed by only slight and gradual changes in the body temperature.

Since the tissues consist largely of water, which has a high thermal capacity and is also a relatively good conductor of heat, the internal parts of the body have a fairly uniform temperature. In addition the circulation helps to distribute heat evenly throughout the tissues. The temperature of the skin is governed to some extent by the external temperature and may fall as low as 17° C if the air is cold. There is thus a considerable temperature gradient in the superficial half-inch or so of tissue made up chiefly of subcutaneous fat and skin (Table 13, 5). Because of this gradient the temperature in the rectum is the best guide to the internal body temperature. Even the rectal temperature, however, may be reduced as the result of cool blood returning along neighbouring veins from cold parts of the surface of the body. The temperature in the mouth is usually 0·5° C lower than the temperature recorded in the rectum but is itself subject to variations, the result, for example, of mouth breathing or the drinking of warm fluids. All these difficulties are not, however, serious enough to destroy the value of records of oral or rectal temperatures in clinical medicine.

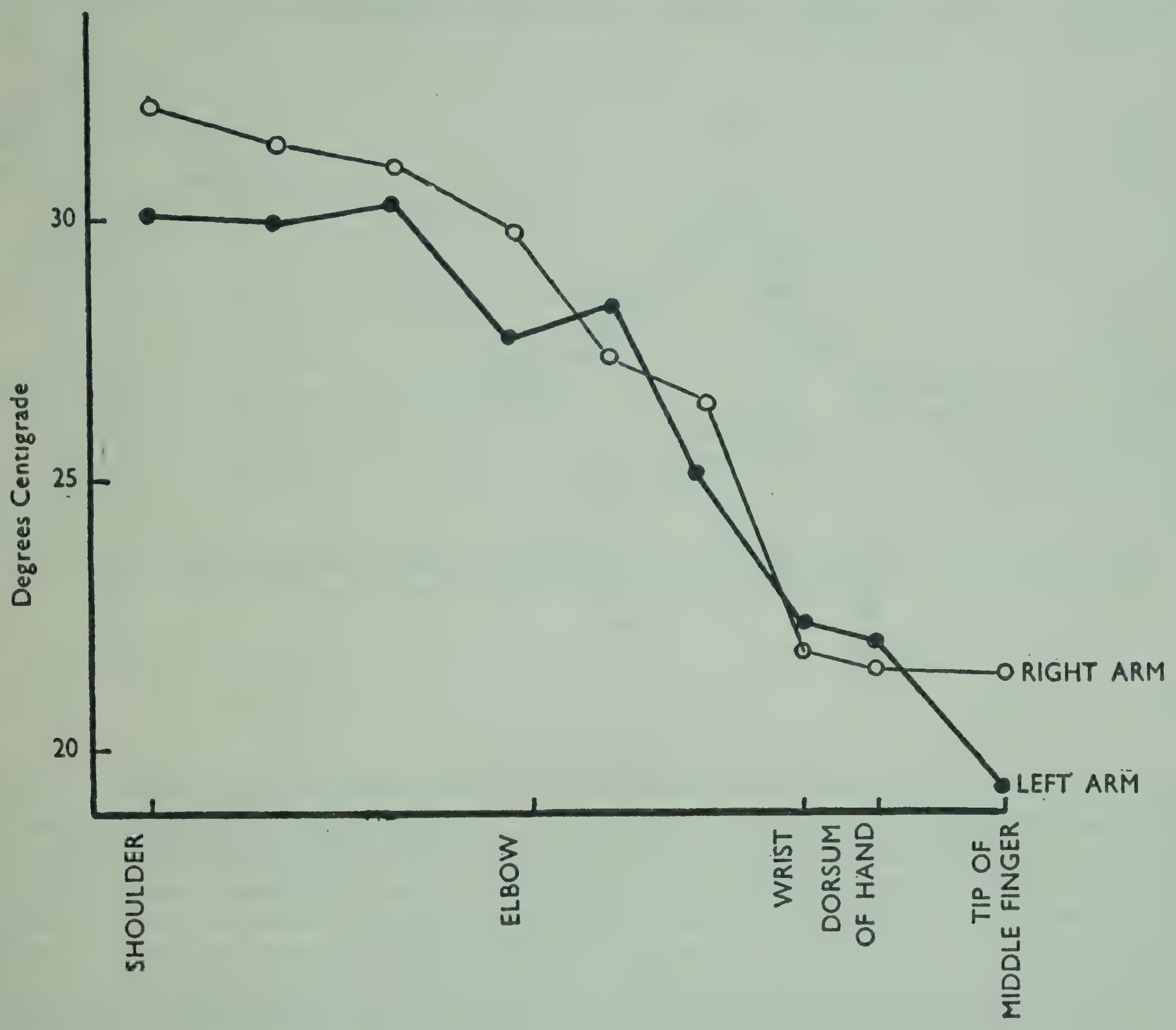


Fig. 13, 2

Readings of skin temperature on the two arms of a healthy subject 30 minutes after the limbs had been exposed to an ambient temperature of 18° C.

The temperature of the skin of the body is by no means uniform. On a limb exposed to an environmental temperature of 18° C the surface temperature of the proximal part of the limb remains higher than that of the distal. This temperature gradient is illustrated in Fig. 13, 2, but the shape of the graph is determined largely by the degree and site of vasodilatation, the distal tissues becoming warmer the more rapid the blood flow through them. It should be emphasized that the temperature of a single organ or region cannot indicate accurately, except by chance, the average internal temperature of the body, since the temperature recorded varies from time to time according to such factors as metabolic activity, the temperature and rate of flow of blood through the region, as well as the heat gradient between the region and surrounding parts. The temperature of the surface of the liver in dogs can be as much as 1·0° C higher than the temperature of the arterial blood, and the temperature of blood in the radial arteries may be as low as 31° C. This low temperature is due to 'pre-cooling' of arterial blood by cool venous blood returning in venæ comites from cold peripheral tissues.

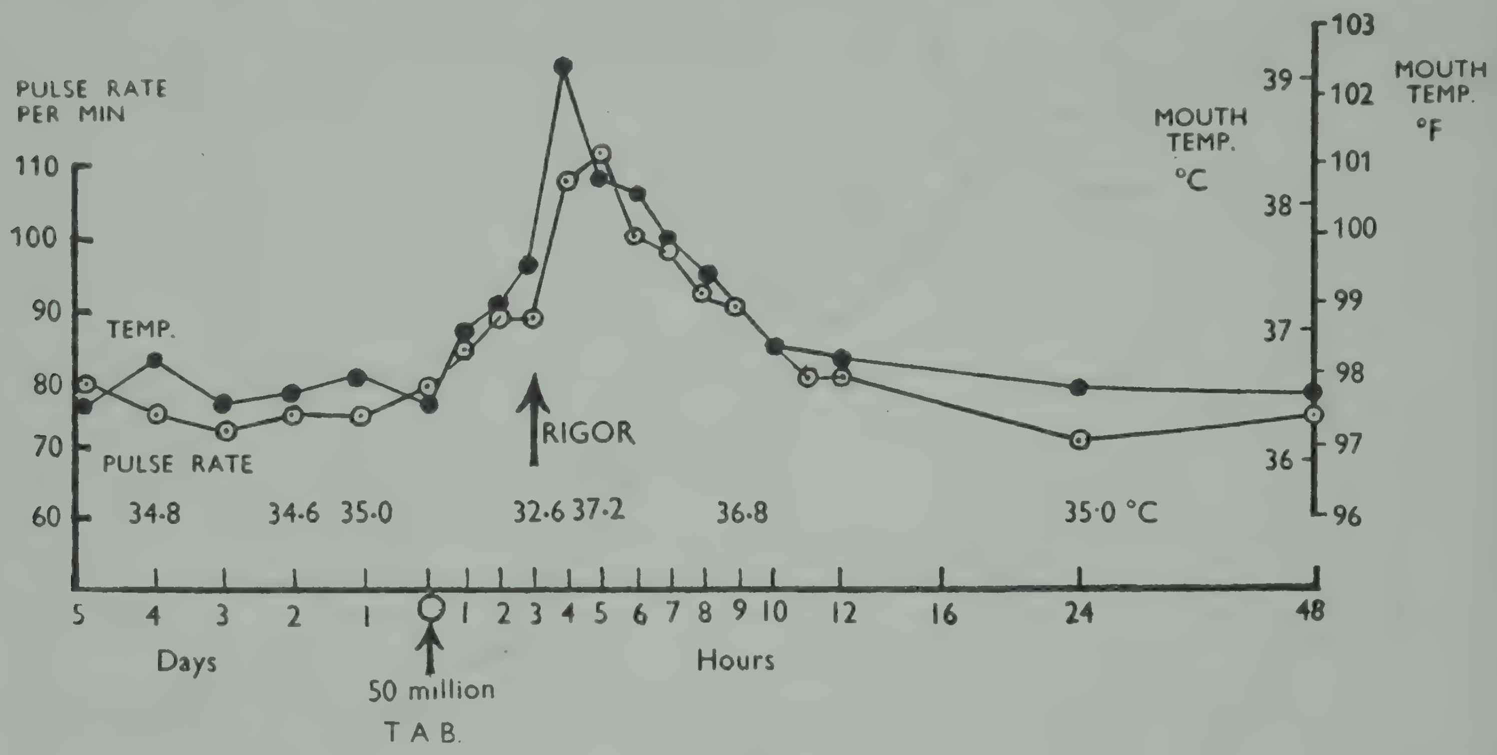


FIG. 13, 3

Chart of mouth temperature and pulse rate in a male patient aged 37 who received an intravenous injection of typhoid vaccine (T.A.B.) containing 50 million organisms. An acute febrile reaction developed three hours later. The figures in the lower part of the chart indicate the temperature of the skin of the forearm measured electrically. At three hours after the injection the patient vomited and complained of pain in the back and legs and headache. He was pale and cold (skin temperature 32·6°). Fifteen minutes later he developed a rigor. His mouth temperature increased rapidly and his skin became flushed and hot (37·2°). Nine hours after the injection the patient felt much better. He was sweating profusely, his mouth temperature had fallen, but his skin temperature remained high.

Heat is produced in the body mainly by oxidations, and the chief sources of heat are therefore the most active tissues, the liver, the glands and the muscles, which make up more than half of the weight of the body. Sudden chilling of the body leads to shivering, which is a general increase in muscular activity by which heat production is considerably raised. Such generalized shivering, called a rigor, may also occur in the initial stages of a fever and may increase the heat production as much as fivefold. During a rigor the peripheral arteries and the skin capillaries are constricted and the patient feels cold although his blood temperature may be rising (Fig. 13, 3). Shivering, the occurrence of which is not closely related to rectal temperature, is probably called forth by nerve impulses from the skin since it can occur with a normal, falling, or rising body temperature and is often initiated, when the body temperature is low, by painful stimuli or when cold air or cold water plays on the skin.

The metabolic rate can be increased considerably by an injection of adrenaline and it may be that sudden exposure to cold is followed by a discharge of adrenaline from the suprarenal glands. This would account for the rise of blood pressure and constriction of the minute vessels which follows sudden exposure to cold. However that may be, adrenalectomy in animals is followed by a fall of body temperature and in chronic adrenal cortical insufficiency in man (Addison's disease) the body temperature is usually low.

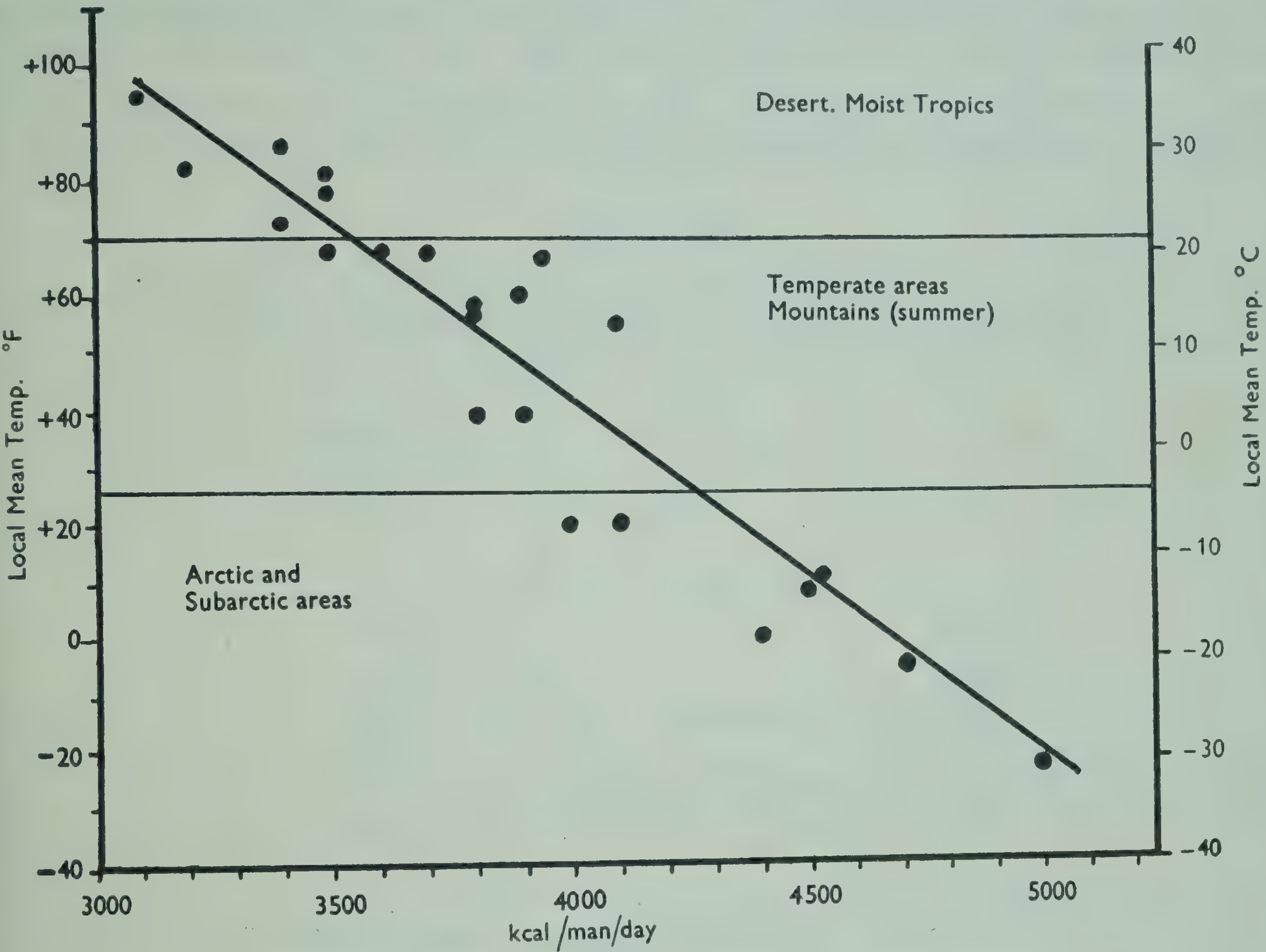


FIG. 13, 4

Voluntary caloric intake, North American troops (averages for groups of 50 or more men). The diagram shows the relation between caloric intake and climatic conditions. The increased requirements in arctic conditions are partly caused by the impeding effect of heavy clothing (R. E. Johnson & R. M. Kark (1947), *Science*, **105**, 378).

As might be expected the body temperature is raised when excessive amounts of thyroid hormone are produced with, consequently, an increased metabolic rate. Conversely diminished thyroid activity is associated with reduced metabolism and subnormal body temperatures. Very low body temperatures are often a feature of hypofunction of the adenohypophysis (Chapter 53). However, in spite of these relationships between body temperature and the thyroid, adrenal and pituitary it is doubtful if the endocrine glands play any important role in the day to day control of body temperature.

In hot climates, with environmental temperatures above 30° C, the basal metabolic rate is usually rather low and the inhabitants tend to be listless and prefer to rest during the hottest parts of the day. In these ways their heat production is reduced. Dwellers in cold climates tend to produce more heat, since their basal metabolic rate is usually a little higher than in the tropics and they are generally more energetic.

In temperate climates (15 to 30° C) the body temperature is not regulated chemically, that is, by varying the heat production, which remains fairly constant. Instead the heat loss is controlled, a procedure which allows of more precise regulation.

The amount of heat produced by a man in a temperate climate depends on the kind and amount of food eaten and metabolized, variations in heat loss compensating for variations in the amount of heat produced at different levels of food intake. In very hot or very cold environments, however, there is a notable alteration in the appetite and in the amount of food consumed, the increased appetite during a cold day in winter being, for example, a familiar experience. Observations made during the 1939–45 war showed that the amount of food eaten by troops with access to as much food as they wanted varied with the local temperature (Fig. 13, 4) but the percentage of kilocalories provided by protein, fat and carbohydrate varied very little with the climatic conditions; that derived from protein, about 12 per cent., was surprisingly constant.

TABLE 13, 5

Temperatures of Different Tissues of the Forearm determined electrically as rapidly as possible after removing Clothes *

	Average Temperature °C
Room	18·5
Skin	33·0
Subcutaneous tissue . .	33·6
Deep muscle . . .	36·2
Mouth (representing arterial temperature) . . .	36·9

* H. Barcroft & O. Edholm (1946), *J. Physiol.* **104,** 366.

In temperate climates the body temperature is nearly always higher than the environmental temperature, so that there is a continuous loss of heat. The fact that there is a temperature gradient between the temperature of the blood and that of the skin (Table 13, 5) shows that the human body is insulated to a certain extent from its environment. Many animals, especially aquatic animals inhabiting cold regions, are insulated by a very thick layer of fat in addition to a coat of fur. The thick layer of insulation provided by the superficial fat in obese persons may so limit heat loss that on a hot day they may suffer much more discomfort than thin persons.

Heat is lost from the body by convection, radiation and conduction. When the air is still, the layer of warm air round the body acts as an insulator because air is a poor conductor of heat. If a wind plays on the naked body this air is continuously being replaced and heat is lost by convection. In cold weather we diminish this loss by putting on clothes which imprison air in their meshes. The amount of clothing is adjusted by experience so that the regulation of body temperature is brought within the range of the physiological mechanisms. The thermal insulation of an air space increases with its thickness up to a maximum of about half an inch, but if the space is filled up with textile material to prevent convection currents the insulation continues to be proportional to the thickness (about 4 Clo units per inch of thickness). For textiles of low bulk-density the insulation provided is the same whether the fabric is wool, cotton, kapok, or any other material. The essential insulator, then, is the air, the particular textile employed to imprison the air making practically no difference to its efficiency as an insulator.

Some 130 kilocalories of the total heat produced are required to warm the food and air taken in. Another 370 kilocalories are lost in evaporating water in the lungs to saturate the expired air, and in liberating carbon dioxide from solution in water. These 500 kilocalories are only 17 per cent. of the total heat output; the remaining 83 per cent. must be lost through the skin. In the basal state there is a loss of body weight, due to evaporation from the lungs and to insensible perspiration, of about 30 g. per hour. This water passes through the skin by diffusion and osmosis, and not as the result of the activity of the sweat glands. Since the latent heat of vaporization of water at body temperature is 0·58 kcal/g. this water loss represents

$$30 \times 0{\cdot}58 \text{ kcal/hr} = 17{\cdot}4 \text{ kcal/hr} = 418 \text{ kcal/day}.$$

For a man of 1·8 sq. metres surface area with a basal metabolic rate of 40 kcal/sq. metre/hour the basal metabolism is 72 kcal/hr (1728 kcal/day). Thus about one-quarter of the heat is lost by the skin even when no sweating occurs. It is usually reckoned that one-quarter of the heat loss occurs in this way whatever the level of activity.

In very hot climates, when atmospheric temperatures exceed body temperature, heat loss must continue although now radiation and convection supply heat to the body. If sweat is secreted and evaporated a considerable amount of heat is taken from the body; if sweat merely drips off, the body does not get rid of heat. The removal of 3000 kilocalories per day, for example, involves the evaporation of $3000 \div 0{\cdot}58 = 5170$ g. of water per day. If this rate of sweating is maintained, then obviously a somewhat greater amount of water must be drunk to provide not only for sweating and saturation of the inspired air but also for kidney and bowel function. Although 5 litres may seem to those in temperate climates a surprisingly large quantity, it is known that as much as 3 litres of sweat may be produced per hour under extreme conditions. Evaporation of sweat occurs readily in a dry atmosphere but with difficulty in a humid one. Thus very hot humid environments are incompatible with life.

In very hot climates the body is heated by the solar radiation which in a tropical desert near noon may amount to 13 kcal/sq. metre/minute. After allowing for the area exposed and for reflexion, the heat absorbed by the body in these circumstances is about 4 kcal/min or 240 kcal/hr, which is more than twice the resting metabolism. A man marching in the desert at 3 m.p.h. produces, say, 270 kcal/hr (3 Met, see p. 209). The total heat to be lost is therefore 510 kcal per hour which requires the evaporation of 880 g. of water per hour.

In such circumstances a man may lose up to 11 litres of water per day, of which 90 per cent. is sweat. If water is not drunk dehydration continues up to 20 per cent. of body weight, which is the limit for survival. Water is not lost equally from all tissues for the blood serum becomes twice as concentrated as the body as a whole. The reduction of circulating blood volume and the increase in viscosity of the blood augments the strain on the circulatory system.

In hot environments the efficiency of mental activity, as tested by accuracy of reception in wireless telegraphy, is reduced. Restful sleep is frequently not obtained in high environmental temperatures and this is probably a factor in reducing the efficiency of work.

SWEAT

Sweat is a watery fluid of very variable composition, but always hypotonic with respect to blood plasma. It contains only about 0·5 per cent. of solids, of which the chief is sodium chloride (0·3 per cent.). It usually contains about 0·03 per cent. of urea, a concentration similar to that in blood, and about 0·07 per cent. of lactate which is produced by the sweat glands. It is remarkable that as the rate of secretion of sweat increases, the concentration of nitrogen in it falls, but the concentration of sodium and potassium actually increases. The concentrations of these ions, and of chloride, in sweat differ from their concentrations in extracellular fluid and experimental evidence suggests that the concentrations in sweat are controlled by the adrenal cortex. (See also p. 99).

The activity of the sweat glands may be studied by measuring the electrical resistance of the skin which becomes progressively reduced as sweating occurs, or alternatively an area of skin is painted with a 3 per cent. solution of iodine in 95 per cent. alcohol and a piece of absorbent paper impregnated with starch then applied. Each active sweat gland shows up a tiny dark blue dot on the paper; the number of dots can be counted with a lens and the sizes of the dots give a measure of the rate of secretion of individual glands.

About 50 per cent. of the total sweat produced on a hot day comes from the skin of the trunk, about 25 per cent. from the lower limbs, and the remaining 25 per cent. from the head and upper limbs. When sweating becomes profuse, so as to give rise to visible drops, it usually does so quite suddenly and more or less simultaneously all over the body. Sweating of the hands and feet is, however, readily produced by anxiety and emotional stress and sweating in these areas should not therefore be regarded as part of the heat-regulating mechanism.

There are in fact two quite different types of sweating, each evoked by a different stimulus. Heating the body leads to profuse sweating especially of the forehead, upper lip, neck, chest and trunk, but to a much less extent of the palms of the hands or the soles of the feet. On the other hand, mental, emotional or sensory stimuli produce sweating mainly confined to the palms, soles and axillæ. Both types of sweating are under nervous control, the temperature-regulating system from centres in the anterior part of the hypothalamus and the emotional system from centres probably in the cortex. The nervous pathways involved are shown in Fig. 13, 6.

Although the sweat glands, of which there are said to be some $2{\cdot}5 \times 10^6$

on the surface of the body, are supplied by post-ganglionic sympathetic fibres the humoral effector substance is acetylcholine. After an injection of atropine sweating is much reduced. This can often be noticed in a hot operating theatre if the patient has had an injection of atropine before the operation to diminish bronchial secretion. Intravenous and intra-arterial injections of adrenaline have no effect on sweating.

Stokers and others who work in very hot environments frequently suffer from severe cramping pains in the muscles in most active use. This 'stokers' cramp' is prevented or relieved by drinking weak salt solution (0·25 per cent. or more). When sweating is profuse as much as 10 g. of chloride an hour may be lost. If water alone is drunk to replace the fluid lost the concentration of sodium chloride in the blood is lowered, and this is apparently the immediate cause of stokers' cramp. Some 8 to 12 g. of salt is normally consumed with the food and no additional salt is needed except in very hot climates when hard work is undertaken. Lee advises from 7 to 21 g. of extra salt per day according to the severity of the climate and the work.

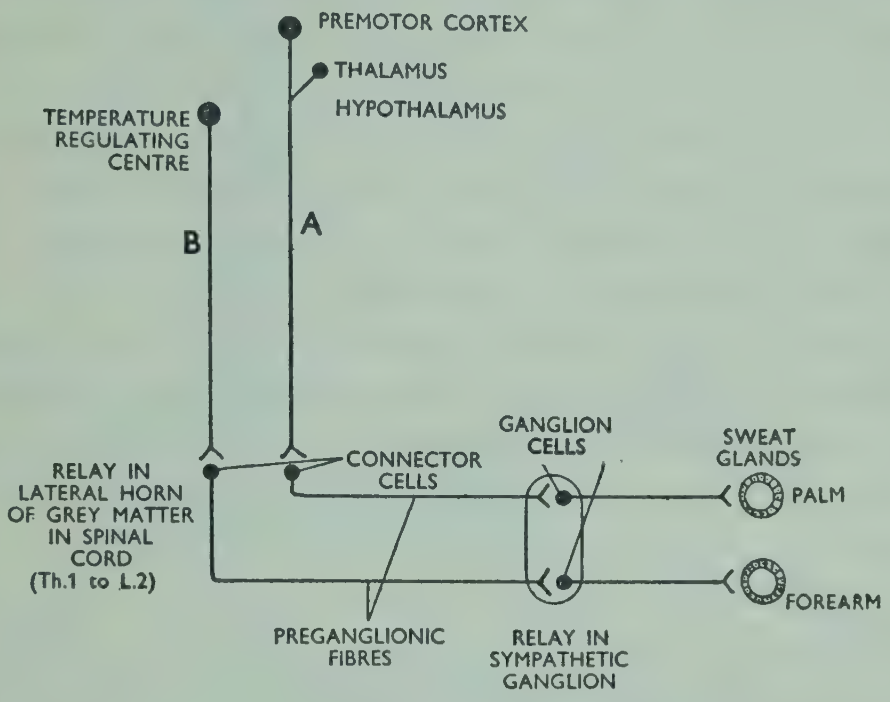

PREMOTOR CORTEX
THALAMUS
HYPOTHALAMUS
TEMPERATURE REGULATING CENTRE
A
B
GANGLION CELLS
SWEAT GLANDS
CONNECTOR CELLS
RELAY IN LATERAL HORN OF GREY MATTER IN SPINAL CORD (Th.1 to L.2)
PALM
FOREARM
PREGANGLIONIC FIBRES
RELAY IN SYMPATHETIC GANGLION

FIG. 13, 6

Nervous pathways of control of sweating. A, pathway from cerebral cortex to sweat glands of palm. These are representative of the glands which take part in 'emotional' sweating. B, pathway from temperature regulating centre in hypothalamus to forearm (T. M. Chalmers & C. A. Keele (1952). The nervous and chemical control of sweating. *Brit. J. Dermatol.* **64,** 43–54).

THE NERVOUS CONTROL OF BODY TEMPERATURE

The nervous system is the most important regulator of the balance between heat production and heat loss. The heat produced by contracting muscles is dependent on impulses arriving from the central nervous system, so that if these are cut off by curare or by a general anæsthetic the body temperature falls. Hence surgical operations under general anæsthesia must be performed in a warm theatre.

Removal of the cerebral cortex in the dog does not seriously impair the control of temperature and it is assumed that this applies also to man. In the paralysis which follows an interruption of the corticospinal fibres in man the skin temperature of the paralysed limb is raised at first, but later falls below that of the unparalysed limb. These disturbances of skin temperature, however, are related to alterations of blood flow rather than to the direct effects of the interruption of nervous pathways. A decerebrate animal which has lost both cortex and thalamus cannot maintain its temperature. The regulating area, lying almost certainly in the hypothalamus, acts on the medullary centres controlling the blood vessels and the secretion of sweat. Vasodilatation and, later, sweating occur when the temperature of the blood in the carotid arteries rises some 0·2 to 0·5° C above normal. If the anterior and middle parts of the hypothalamus are warmed, or if this area is stimulated electrically, the animal sweats and pants, and there is vasodilatation; in other words, it exhibits the reactions seen when the environmental temperature is high. Destructive lesions of this so-called parasympathetic area diminish the animal's ability to withstand high temperatures. Stimulation of the posterior (or sympathetic) area of the hypothalamus causes vasoconstriction; destructive lesions in this area lead to vasodilatation and lowering of temperature without shivering. Shivering, which is a normal response to a lowering of body temperature, seems to have its origin in the hypothalamus, the efferent paths lying in the tectospinal and rubrospinal tracts rather than the corticospinal (pyramidal) tract. There may be a centre in the brain stem just above the level of the pons which when cooled initiates shivering.

Apart from sweating the loss of heat from the skin depends on the amount of warm blood flowing through it. This is chiefly determined by the rate of blood flow through arteriovenous anastomoses with which the skin of the peripheral parts of the limbs is particularly richly equipped. These anastomoses (Fig. 13, 7), which form direct communications some 35 μ in diameter between arteries and veins, are unusually richly supplied with sympathetic nerve fibres. When they are widely open blood is rapidly shunted from the arteries to the veins and the blood flow through the skin is greatly increased. Such anastomoses subserve two main functions. When the part is exposed to cold they serve to maintain its temperature by allowing blood to flow through the skin at a time when the capillaries are constricted. They are also important in the regulation of body temperature, since the loss of heat from the body is greatly facilitated by the enormous increase in blood flow which occurs when these shunts are widely open. The state of the arteriovenous anastomoses, while partly governed by local factors such as axon reflexes, is regulated chiefly by variations in sympathetic nervous tone. When the temperature of the blood passing through the hypothalamic centres is increased, and heat must be lost, sympathetic vasoconstrictor tone is reduced, the arteriovenous anastomoses dilate and warm blood reaches the skin in greater quantity.

In cases of hæmorrhage into the mid-brain or pons the balance between heat production and heat loss is seriously disturbed and the body temperature may be either very high or very low. This is probably due to the isolation of the heat-regulating centres from the blood vessels of the skin where the actual heat loss takes place. In these circumstances the patient behaves like a poikilothermic animal. If he is covered with blankets and surrounded by hot-water bottles his body temperature quickly rises to a dangerous level.

VARIATIONS IN BODY TEMPERATURE

We have already seen that the temperature varies at different parts of the body and also that the temperature of the various organs alters from time to time. Furthermore it has been shown that the temperatures of different internal organs can change simultaneously in opposite directions. In fact "the conception that a man has an internal mass of tissue maintained at 37° C (98·6° F) surrounded by a thin layer with a steep temperature gradient, such that the surface temperature is about 33° C (91·4° F), is common but inaccurate" (Bazett). Even if we simply take the temperature in the mouth (representing blood temperature) of a healthy man at intervals during the day we find that it varies from, for example, 35·8° C (96·5° F) to 37·3° C (99·2° F), being highest in the late afternoon and lowest in the early morning (Fig. 13, 8). Exercise, because it increases production of heat, may elevate the oral temperature 1° or 2° C (Chapter 42) and violent exercise can produce temperatures of 39° C (102·2° F). Oral temperatures in general tend to be higher in children and lower in the elderly. In newborn babies, especially if they are born prematurely, the rectal temperature is normally about 2° F (1·1° C) below the adult value which is usually not

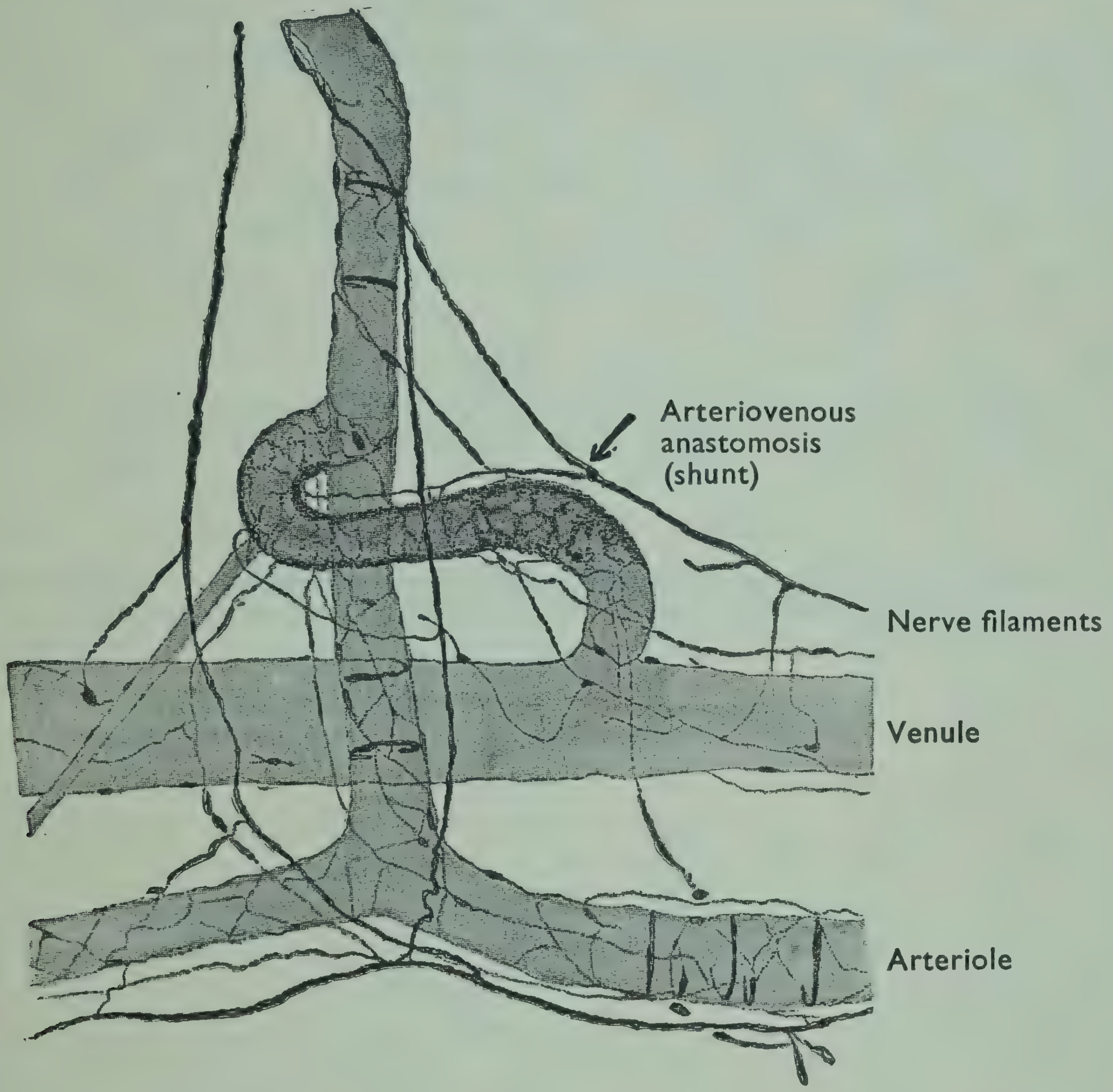


FIG. 13, 7

A projection drawing (× 120) of a single arteriovenous anastomosis in a rabbit's ear. Similar structures can be demonstrated in human skin (R. T. Grant (1930), *Heart,* **15,** 303).

reached until some 10 to 14 days after birth. Thus, in the neo-natal period, a rectal temperature of say 99 to 99·5° F (37·2 to 37·5° C) which might be interpreted as normal, is actually equivalent to a temperature of 101° F (38·3° C) in an adult.

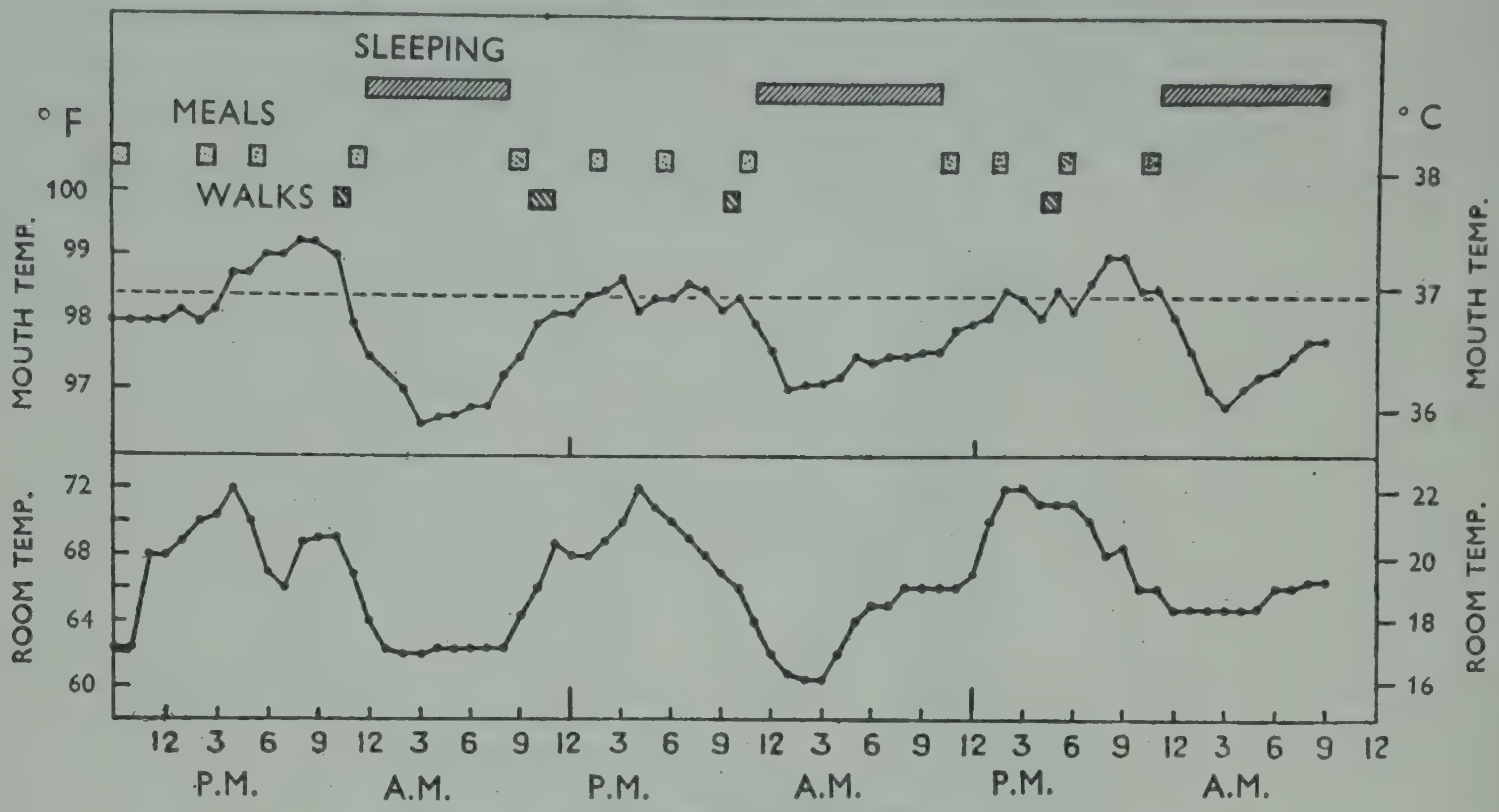


Fig. 13, 8

Diurnal variation of body temperature in a healthy person in relation to external temperature and bodily activities (Samson Wright (1949), *Brit. med. J.* i, 610).

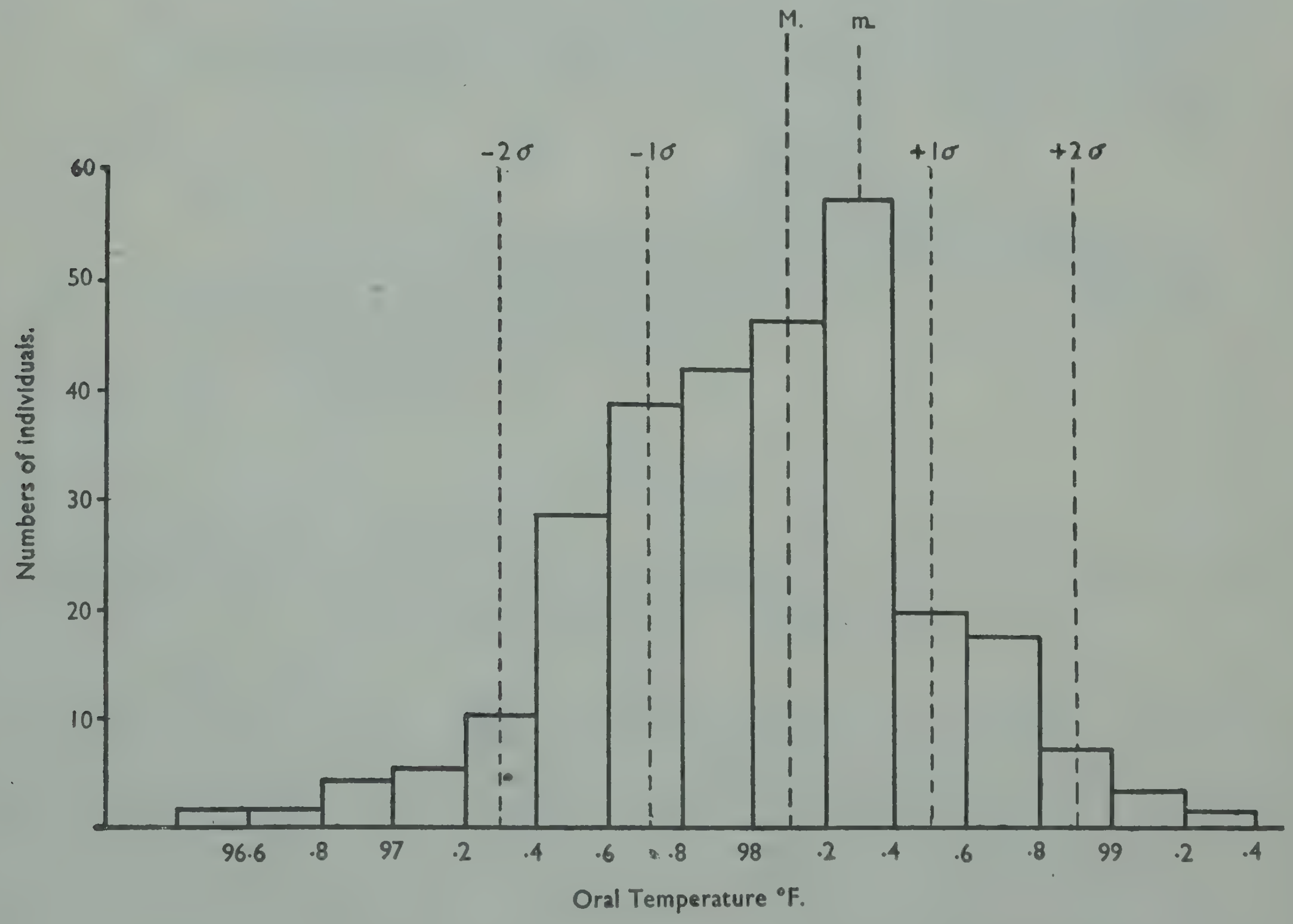


Fig. 13, 9

A histogram showing the frequency distribution of oral temperature in 276 medical students. All temperatures were taken between 8 a.m. and 9 a.m. with the subjects seated in a warm class-room. M = arithmetic mean of the series; m = mode of the series; σ = standard deviation. Data from A. C. Ivy (1944), *Quart. Bull. Northw. Univ. med. Sch.* **18**, 22, taken with modifications from Du Bois (1949). *Fever and the Regulation of Body Temperature.* Springfield, Illinois: Thomas.

When we come to consider the temperatures of a number of healthy persons, even of the same age and sex, recorded under standard conditions and at the same time of day, it is not surprising to find differences between them as shown in Fig. 13, 9. The temperatures ranged from 35·8° C (96·4° F) to 37·3° C (99·2° F). The arithmetic mean was 36·7° C (98·1° F) but to refer to this as 'the normal temperature' and to have a mark made at this level on clinical thermometers would obviously be misleading. The temperature most frequently recorded (the mode), namely 36·8° C (98·2° F), is probably a better index of 'normality' although the difference is of little importance in clinical medicine.

It is apt to be assumed that increase of body temperature (*pyrexia*) is due to an increase of heat production. Although this may occur, an increase in temperature is more often brought about by a diminished heat loss. Even when the body temperature is raised there is still a balance between heat production and heat loss, but at a higher level than normal, as if the hypothalamic thermostat had been 'set' at a new and higher level.

When the regulating mechanisms break down and allow extreme variations of body temperature serious results follow; body temperatures above 43° C (109° F), if they persist for any length of time, are incompatible with life. Even a small elevation of temperature produces an increase in cardiac and respiratory rates, and body temperatures above 40·5° C (105° F) are accompanied by weakness, headache, circulatory collapse and finally loss of consciousness. When the temperature of the body is reduced to about 27° C (80° F), by packing it about with ice or in some other way, then the metabolism is reduced some 30 to 40 per cent. The heart and respiratory rates are much reduced, arterial pressure is very low and the patient passes into a state of suspended animation from which he will completely recover when he is allowed to warm up. In such a state the circulation can be interrupted for short periods of time so that surgical operations on the heart and large blood vessels become practicable. Below 25° C in dogs, cardiac irregularities and ventricular fibrillation tend to develop suddenly and it is no longer possible to resuscitate the animal. Packing ice round a limb can render it anæsthetic and make possible surgical operations such as amputations.

CLOTHING

The functional design of clothing requires a knowledge both of physiology and of physics. Comfort is dependent largely on skin temperature, the optimum being about 33° C. A skin temperature of 33° can be maintained in a cold environment only if the clothing is properly chosen in relation to the activity of the subject.

It has been proposed to use as a convenient unit of metabolic activity the Met, that is the metabolic rate of a man at rest.

0·8 Met	40 kcal/sq.m/hr	B.M.R.
1 Met	50 kcal/sq.m/hr	metabolic rate at rest
3 Met	150 kcal/sq.m/hr	,, ,, walking
6 Met	300 kcal/sq.m/hr	,, ,, running

Thermal insulation can be considered as the resistance to flow of heat and can be calculated, like electrical resistance, by a simple formula similar to that derived from Ohm's law $\left(R=\frac{E}{C}\right)$ thus:

$$\text{Thermal insulation}=\frac{\text{Driving force (that is difference of temp.)}}{\text{Flow of heat per unit of cross sectional area}}.$$

The unit adopted for measuring the thermal insulation of clothing, the Clo, is taken as the insulation needed to keep a resting sitting subject comfortable in still air at 21° C. If it is assumed (p. 203) that about one-quarter of his heat loss occurs by insensible perspiration then $\frac{3}{4}$ Met or 38 kcal/sq.m/hr is transmitted through the clothing. If I_{CL} and I_A are respectively the insulation of the clothing and of the layer of air around the clothing the total resistance

$$I=I_{CL}+I_A=\frac{33-21^\circ\text{ C}}{38\text{ kcal/sq.m/hr}}=0\cdot32^\circ\text{ C/kcal/sq.m/hr}$$
$$=0\cdot32\text{ metric units.}$$

The insulation value of still air I_A at 21° is by experiment 0·14 metric units, so that

$$I_{CL}=0\cdot32-0\cdot14=0\cdot18\text{ metric units.}$$

This value of resistance is one Clo by the definition just given. One Clo represents the insulation provided by ordinary everyday clothing about 6 mm ($\frac{1}{4}$ in.) thick. Putting on an overcoat increases the insulation to 2 Clo, but as much as 4 Clo can be provided by special clothing.

TABLE 13, 10

Variations in Insulation of Ambient Air *

I_A *in Clo Units*

Air Movement	Resting Sitting	Level Walking		
		Slow	Normal	Fast
Normal indoors (20 ft./min) . .	0·78	0·37	0·33	0·29
Draughty indoors (100 ft./min) .	0·40	0·32	0·29	0·26
Normal outdoors (5 m.p.h.) . .	0·30	0·24	0·22	0·20
Windy outdoors (20 m.p.h.) . .	0·19	0·18	0·17	0·16

* From A. P. Gagge, A. C. Burton & H. C. Bazett (1941), *Science*, **94**, 428.

The insulation provided by the air round the body depends on the rate of air movement, and will obviously be reduced if air movement is increased by wind or the activity of the subject. Values obtained by experiment are given in Table 13, 10. The insulation varies from about 0·8 Clo (0·14 metric units) downwards according to circumstances.

Since the skin temperature is lower than blood temperature the superficial tissues offer resistance to outflow of heat which in the example given above may be calculated as follows:

$$\text{Heat resistance of tissues} = \frac{37-33}{50} = 0{\cdot}08 \text{ metric units, or } \frac{0{\cdot}08}{0{\cdot}18} = 0{\cdot}44 \text{ Clo units.}$$

The insulation resistance of the superficial inch of the body is about the same as that of wood and half as much as that of flannel, so that the superficial tissues of the body are surprisingly good insulators in spite of the blood circulating in them.

From the information just given the clothing required in various cold climates for different grades of activity can be calculated. For example, the amount of clothing required for comfort at – 1° C with normal outdoor conditions when walking at a normal pace on the level is calculated as follows. Level walking at a normal pace produces 3 Met (see p. 209), one-quarter of which is lost by evaporation, leaving three-quarters to be lost through the clothes. By substituting these values in the equation on p. 210 the total insulation I is obtained thus

$$\begin{aligned} I &= \frac{33-(-1)}{\frac{3}{4}\times 3\times 50} \text{ metric units} \\ &= \frac{34}{\frac{3}{4}\times 3\times 50\times 0{\cdot}18} \text{ Clo units} \\ &= 1{\cdot}68 \text{ Clo units.} \end{aligned}$$

But 0·22 Clo units are provided by the layer of air (Table 13, 10), so that 1·46 Clo units of clothing are required. The results of a number of such calculations are given in Table 13, 11.

TABLE 13, 11

Ideal Clothing for Comfort in Clo units

Environmental Temperature *	Resting Sitting 1 Met	Slow Level Walking 2 Met	Normal Level Walking 3 Met	Fast Level Walking 4 Met
21° C (70° F)	1·5	0·7	0·4	0·3
10° C (50° F)	3·1	1·5	0·9	0·7
– 1° C (30° F)	4·7	2·3	1·5	1·1
–17° C (0° F)	7·2	3·5	2·3	1·7

* Air movement normal outdoors (5 m.p.h.). From A. P. Gagge, A. C. Burton & H. C. Bazett.

In cold environments it is extremely important to keep the hands warm if delicate work has to be carried out. The problem cannot be solved by insulation alone, since gloves beyond a certain thickness are exceedingly clumsy. This difficulty has been overcome by incorporating electrical heating in airmen's gloves and clothes. Although the hands and feet feel cold first, it is necessary in extreme conditions to provide heat for the whole body as well as the fingers. Even when the physical requirements for the design of clothing for very cold environments have been fulfilled certain physiological difficulties persist. In normal circumstances shivering occurs before the rectal temperature has fallen more than 0·1° C or 0·2° C, but a heavily clothed man in a very cold environment may suffer a drop of rectal temperature of 1° C without shivering. Because of the discomfort which accompanies his persistently low body temperature his mental and physical performance is reduced.

If the insulating properties of the clothing are to be maintained the air in its meshes must be dry. People who work in Arctic conditions are agreed that the accumulation of moisture from the body interferes with efficient insulation more than moisture derived from rain or snow. If sweat condenses on the cold outer layer of the clothing the thermal insulation is much reduced. Water-repellent, but not waterproof, garments are required in these conditions. In shipwreck also it is important to try to start with dry clothes and to keep them dry. Wet clothing may contain, say, 3 kg. of water, the evaporation of which at body temperature requires

$$3000 \times 0{\cdot}58 = 1740 \text{ kcal}$$

which is the equivalent of a good emergency food ration for a day. Water-repellent exposure suits have been specially designed so that they can be continuously worn over the clothing. They prevent spray from soaking the clothing but at the same time allow water vapour derived by evaporation from the body to escape. Low resistance to transference of water vapour is also an important property of clothing used in hot climates. Closely woven cloth is much more suitable in this respect than thicker material with an open weave.

Shipwrecked men immersed in very cold water lose heat rapidly, and if they try to save their lives by clinging to wreckage or by floating on their lifebelts they quickly perish. On the other hand vigorous swimming, even in water at 0° C, can produce nearly as much heat as is lost and the survival time is greatly prolonged by the exercise.

Measurement of Environmental Conditions

The ways by which man can adapt himself to changes in environmental temperature have been described. But nowadays the converse process by which man makes his environment suit his comfort is frequently applied. Unfortunately the reactions of a human being to different climatic conditions cannot easily be assessed and cannot be measured. They depend on individual ideas of comfort and discomfort, as well as on the physical factors of temperature, humidity and rate of air movement. It is not surprising therefore to find that no simple and yet satisfactory method of estimating the suitability or otherwise of an environment either indoors or out-of-doors has been devised. Knowledge of air temperature itself is of little value, because a warm humid climate may be uncomfortable when a warm dry one is quite comfortable. The humidity is most readily calculated from the readings of the wet and dry bulb thermometers. A wet bulb temperature of 30° C (85° F) may be tolerable at rest in still air, but even in mild exercise the maximum tolerable wet bulb temperature may be 25° C (78° F). The importance of the rate of air flow is illustrated by the well-known fact that the feeling of stuffiness in a hot still environment can be dispelled if the air is stirred by a fan. If the air velocity and wet and dry bulb readings are known a fair estimate of the environment can be made by using the charts provided by Bedford (see reference below).

HEAT STAGNATION.—The cause of stuffiness in a room full of people was not understood until relatively recently. It was first supposed that the cause of the discomfort was the reduction in oxygen content of the air. In point of fact, however, a room may become very stuffy long before the oxygen percentage is reduced to 20 per cent., whereas fresh air containing 16 per cent. oxygen is quite tolerable. Again, the carbon dioxide content, even in badly ventilated rooms, never rises above 1 per cent., an amount which, in fresh air, would scarcely be noticed. Thus chemical theories of the vitiation of the atmosphere had to be abandoned. Later hypotheses that the cause was a respiratory poison or a bacterial infection were also discarded, because of the rapid onset of symptoms and the quick recovery when the sufferer was brought into fresh air. J. S. Haldane in 1905 was the first to arrive at the true explanation which is that the symptoms are really due to heat stagnation. In a hot, moist, still environment heat loss is profoundly impaired and these are the conditions which led to the death of 123 out of 146 British prisoners confined overnight in the Black Hole of Calcutta, a place eighteen feet by eighteen feet by nine feet. If we assume, for purposes of calculation, that no victim was able to lose any heat, the rise of body temperature which would have occurred can be reckoned from the water equivalent and the heat loss. If the specific heat of the body is 0·8, the water equivalent of a 70 kg. man is 56 litres. If his metabolic rate was 72 kcal per hour, his body temperature would rise $72 \div 56 = 1{\cdot}3°$ C per hour, and death would occur in three or four hours from hyperpyrexia.

REFERENCES

ADOLPH, E. F. & Associates (1947). *Physiology of Man in the Desert.* London: Interscience.

BAZETT, H. C. (1927). Physiological responses to heat. *Physiol. Rev.* **7**, 531–599.

BEDFORD, T. (1946). Environmental warmth and its measurement. *M.R.C.* War Memo. No. 17. London: H.M.S.O.

BURTON, A. C. & EDHOLM, O. (1954). Man in a cold environment. *Physiol. Soc. Monograph* No. 2. London: Arnold.

CHALMERS, T. M. & KEELE, C. A. (1952). The nervous and chemical control of sweating. *Brit. J. Dermatol.* **64**, 43–54.

CROWDEN, G. P. (1949). A survey of physiological studies of mental and physical work in hot and humid environments. *Trans. R. Soc. trop. Med. Hyg.* **42**, 325–335.

DU BOIS, E. F. (1937). The mechanism of heat loss and temperature regulation. *Lane Lectures.* Stanford University Press.

DU BOIS, E. F. (1949). *Fever and the Regulation of Body Temperature.* Springfield, Illinois: Thomas.

GARRY, R. C., PASSMORE, R., WARNOCK, G. M. & DURNIN, J. V. G. A. (1952). Studies on expenditure of energy and consumption of food by miners and clerks, Fife, Scotland, 1952. *Spec. Rep. Ser. med. Res. Coun., Lond., No.* 289. London: H.M.S.O.

KUNO, Y. (1934). *The Physiology of Human Perspiration.* London: Churchill.

LEE, D. H. K. & LEMONS, H. (1949). Clothing for global man. *Geogr. Rev.* **39**, 181–213.

NEWBURGH, M. D. (Ed.) (1949). *Physiology of Heat Regulation and the Science of Clothing.* London: Saunders.

ROBINSON, S. & ROBINSON, A. H. (1954). Chemical composition of sweat. *Physiol. Rev.* **34**, 202–220.

CHAPTER 14

HUNGER AND THIRST

IN the alimentary canal the food is broken up and mixed mechanically with the digestive juices which hydrolyse the complex molecules of the foodstuffs to relatively simple molecules that can be absorbed. The products of digestion are absorbed into the body through the walls of the canal; a certain amount of material is neither digested nor absorbed and is stored for a time before being rejected. The lower parts of the alimentary canal contain enormous numbers of bacteria which play a significant role in nutrition, especially in the synthesis of vitamins. The bacteria, although resident in the intestine, are not within the tissues and they do not give rise to pathological changes in the wall of the intestine unless the gut is damaged. Man has voluntary control over the intake of food through the mouth and pharynx and the discharge of material from the anus but the other activities occur automatically and, for the most part, without his being aware of them.

The sensations of hunger and thirst, of prime importance to the survival of the individual, are influenced by the functioning of the alimentary canal and they must be considered now.

HUNGER

Hunger is a complex sensation partly somatic and partly psychological in origin. There is considerable evidence that a feeling of hunger is often associated with gastric contractions but it can occur even after all the stomach has been removed. Cannon, who was the first man to use opaque meals for X-ray examination of the gut, found that hunger was often associated with slow tonic gastric contractions each of which raised the intragastric pressure. Carlson and also Wolf and Wolff in their studies of men with gastric fistulæ found that a phase of increased gastric motility occurred every two or three hours when the stomach was empty. The whole phase, which consisted of a series of contractions each lasting about twenty seconds, had a duration of ten to thirty minutes; only fifty per cent. of these phases, however, were accompanied by a sensation of hunger.

The contractions of the empty stomach are inhibited by swallowing or chewing or smoking, by exercise or exposure to cold, by intravenous injection of glucose or by taking fluid by mouth (p. 233). Such contractions are intensified by the sight and smell of food. Persons fasting for experimental purposes have reported that they became hungry at their customary meal times. It may be that there is some association between contractions of the stomach and the sensation of hunger, but the relationship is certainly not close enough to allow of the definite statement that strong gastric contractions are always the somatic cause of hunger. Indeed after gastrectomy in man hunger is experienced in much the same way as in other people. Dogs after denervation of the stomach and upper intestine by vagotomy and sympathectomy still show evidence of hunger although no gastric contractions occur. Furthermore the drug amphetamine abolishes the hunger of these dogs just as it does in intact animals.

When the blood sugar is lowered to, say, 40 mg. per 100 ml. by injection of insulin, gastric secretion and contractions are increased and the patient may feel weak and 'empty.' It is doubtful, however, if this sensation can be described as hunger and, in any case, as the following examples show, there is no constant relationship between blood sugar and the feeling of hunger. In healthy hungry fasting persons the blood sugar is within normal limits, that is between 70 and 90 mg. per 100 ml. Hunger may be a prominent symptom of untreated diabetes mellitus with a high blood sugar. Patients with a continuous low level of blood sugar (for example as the result of hyperinsulinism) do not always complain of constant hunger.

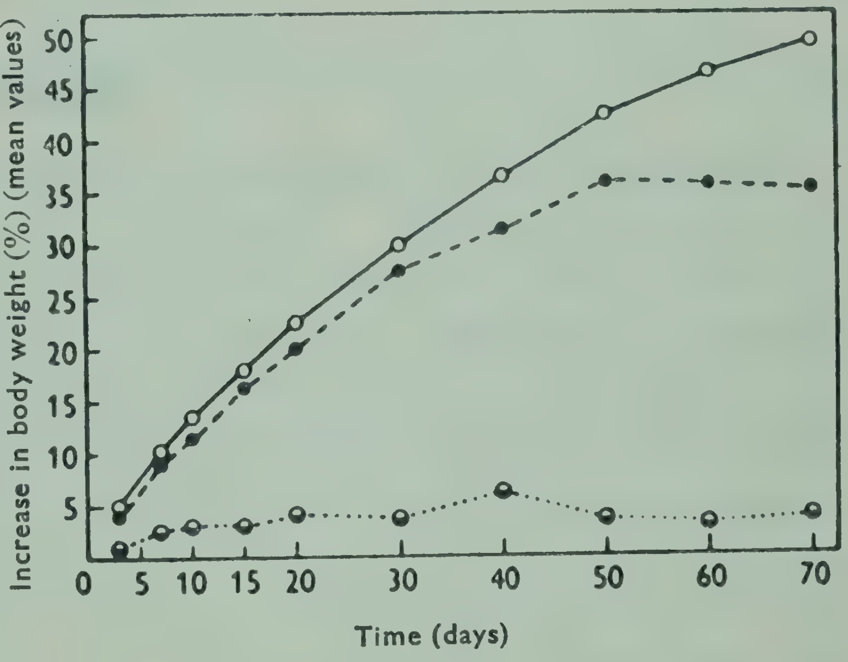


FIG. 14, 1

Growth curve ○—○ of adult rats injected with growth hormone and given unlimited access to food. Growth curve ◒.....◒ of control rats and growth curve ●- - - -● of hormone-treated rats given the same quantity of food as the control group (A. L. Greenbaum (1953) *Biochem. J.* **54,** 401).

There is evidence for a hunger centre in the brain; bilateral damage to the ventromedial nuclei of the tuber cinereum of the hypothalamus causes obesity in rats but bilateral lesions placed more laterally abolish hunger and the animals die of starvation. Although the hypothalamus is commonly regarded as the site of the hunger centre there can be no doubt that lesions of the frontal cortex may also cause increased consumption of food.

The terms 'hunger' and 'appetite' are often used indiscriminately but physiologists tend to assign separate meanings to them. Hunger is usually taken to mean the sensation of emptiness resulting from abstinence from food, that is an unconditioned phenomenon. On the other hand appetite is thought of as the desire for food or the pleasure in taking food which is conditioned, or learned. Thus hunger may be thought of as a physical or physiological phenomenon and appetite as a psychological one. Sometimes appetite acquires a specific character as in the case of wild animals which travel long distances to obtain salt. In adrenal insufficiency in both man and animals there is an increased desire for sodium chloride.

Since body weight in adult life remains relatively constant one is forced to conclude that hunger and appetite are governed by the needs of the body so that a balance between food intake and energy output is achieved. It is, therefore, appropriate to enquire not only what it is that makes people begin to eat but also what makes them stop. We have no clue as to the signal which makes us stop eating but observations show that we tend to be satisfied when the weight of the daily food intake reaches 5 per cent. of the body weight.

The balance between body needs and hunger has long been suspected to be under hormonal influence since, to quote one example, castrates tend to become rather fat. A clue to this influence may be furnished by the experiments of Greenbaum (Fig. 14, 1). Adult rats injected with growth hormone (Chapter 53) while obtaining access to unlimited supplies of food increased in weight up to 50 per cent. in 70 days. A second group of rats also injected with growth hormone, but restricted to the dietary intake of a third control uninjected group, increased their weight by 28 per cent. in 30 days much as did the first group; after 30 days the gain in weight of the second group slowed up. In other words, the animals injected with growth hormone made a different use of the same quantity of food; they converted a certain amount of it to body tissue whereas the uninjected controls (the third group) oxidized more and did not gain in weight.

THIRST

Shortage of water causes more immediate and more intolerable distress than shortage of food. A person completely deprived of water soon feels his mouth dry, he complains of thirst and the craving for fluid rapidly becomes compelling. As time goes on the dryness of the mouth increases, the production of saliva ceases and the swallowing of food becomes impossible. Finally delirium is followed by death within a day or two in a dry climate or a little longer in a moist environment.

Normally the intake of water in food and drink is so regulated that it balances the loss of water in the urine, faeces, sweat and breath. The body weight, and the amount of water in the body, thus remains constant. In exceptional circumstances this may not be quite true since men, actively working during the day in the desert and sweating profusely, drink only a half of the water they are losing. The consequent dehydration of the body reduces their physical efficiency. The body weight is regained by an increased fluid intake in the evening.

The sensation of thirst is to be regarded as the expression of a general bodily need for water but the way in which the sensation is produced is not known. It is not the result merely of a dry mouth because, although persons who are thirsty generally have a dry mouth, the converse is certainly not true. Furthermore an injection of atropine, sufficient to stop the secretion of saliva, though producing a dry mouth, does not make a person thirsty and patients who have *xerostomia* as the result of disease do not complain of thirst.

On the other hand, the injection intravenously of hypertonic solutions, for example 5 per cent. sodium chloride, regularly produces thirst. Such solutions increase the osmotic pressure of the extracellular fluid and are, therefore, followed by a movement of fluid out of the cells, the osmolarity of which is relatively increased. When, however, a reduction of cell water is produced without any change in osmolarity, as for example in potassium depletion (Chapter 35), thirst is not produced. It is therefore likely that the sensation of thirst is in some way the result of an increased osmotic pressure of the fluid inside the cells and calculations based on experiments in man with hypertonic saline show that thirst is produced when about 1 per cent. of the intracellular water of the body has passed into the extracellular space. In the thirst that follows the consumption of alcohol there is also an increase in the extracellular fluid at the expense of the intracellular water.

This hypothesis satisfactorily accounts for the fact that thirst occurs naturally when there is 'predominant water depletion' (Chapter 35), that is to say when the intake of water is low relative to the intake of salt, the osmotic pressure of the plasma is increased, the output of anti-diuretic hormone from the pituitary (Chapter 34) is increased and thus the urinary loss of fluid decreased. Although it is true that people who lack water but do not lack salt are thirsty this cannot be the whole explanation because not all persons who complain of thirst are in that state. Furthermore the absence of thirst is not a guarantee that the subject does not require water.

REFERENCES

Harris, S. C., Ivy, A. C. & Searle, L. M. (1947). The mechanism of amphetamine-induced loss of weight. A consideration of the theory of hunger and appetite. *J. Amer. med. Ass.* **134,** 1468–1475.

Hollander, F. (Ed.) (1955). The regulation of hunger and appetite. *Ann. N.Y. Acad. Sci.* **63,** 1–144.

Holmes, J. H. & Montgomery, A. V. (1953). Thirst as a symptom. *Amer. J. med. Sci.* **225,** 281–286.

Janowitz, H. D. & Ivy, A. C. (1949). Role of blood sugar levels in spontaneous and insulin-induced hunger in man. *J. appl. Physiol.* **1,** 643–645.

Sangster, W., Grossman, M. I. & Ivy, A. C. (1949). Relation of gastric evacuation to time of onset of gastric hunger contractions. *J. appl. Physiol.* **1,** 637–642.

Strang, J. M. (1951). Satiety as a factor in nutrition states: observations on mass exchange. *Amer. J. med. Sci.* **221,** 537–542.

Wolf, A. V. (1950). Osmometric analysis of thirst in man and dog. *Amer. J. Physiol.* **161,** 75–86.

CHAPTER 15

MOUTH, ŒSOPHAGUS AND DEGLUTITION

THE oral cavity and tongue are covered by a stratified squamous epithelium richly supplied with pain and tactile nerve endings; in addition the tongue has taste buds distributed over its surface. If, when food enters the mouth, the impulses from these nerve endings and those in the olfactory area indicate that the food is acceptable, mastication and swallowing follow.

MASTICATION

Solid food is broken up in the mouth by chewing which leads at the same time to an increase of the salivary secretion and mixes up the food with the saliva to make a bolus suitable for swallowing. The movements of chewing have been studied by Sherrington in the decerebrate preparation (Chapter 46), in which the jaws are kept closed as part of the characteristic posture in which all the anti-gravity muscles are in action. If the gum near the teeth of this preparation is stimulated mechanically or electrically, or if pressure is applied to the crown of a tooth, the jaw opens reflexly and closes very strongly immediately the stimulus is removed (rebound-contraction). The normal movements of mastication may, therefore, be described as follows. When a piece of food is taken into the mouth the teeth are voluntarily closed on it but the pressure on the teeth causes the jaws to open reflexly. The rebound of the motoneurones (Chapter 45) of the jaw-closing muscles causes a powerful closing, which produces once more a jaw-opening reflex, and the process continues so long as there is anything left in the mouth to chew. The nervous mechanism is probably similar in man and dog, because jaw-opening in the unanæsthetized human subject can be produced by electrical stimulation of the mucosa of the cheek opposite the second upper premolar tooth.

SALIVARY DIGESTION

The Salivary Glands.—In man and in many animals the saliva is secreted by three pairs of glands: the *parotid*, the *submandibular* and the *sublingual*. The duct of the parotid gland enters the mouth opposite the second upper molar tooth. The submandibular gland has a single duct which reaches the mouth at the side of the frenulum of the tongue, but the secretion of the sublingual glands passes by way of some 10 to 20 small ducts which open into the floor of the mouth. There are in addition numerous small salivary glands scattered over the inside of the mouth. In animals saliva may be collected from the parotid or the submandibular gland by means of a cannula placed in the duct, but a more convenient method, used by Pavlov (Chapter 50), is to free the terminal portion of the duct, with the area of mucous membrane around it, and transfer it to the external skin surface to form a salivary fistula. In man parotid saliva can be obtained by inserting a soft plastic cannula into the parotid duct.

The Histological Structure of the Salivary Glands.—The salivary glands are typical compound tubular glands in which the glandular cells lie within a supporting framework of connective tissue carrying the blood vessels, lymphatics and nerves. The glandular cells are arranged in a single layer around a central cavity which receives the secretion from the surrounding cells (Fig. 15, 2). Such a unit is called an *acinus* or an *alveolus*. Small ducts from a number of adjacent acini join together to form larger ducts, and the secretion from the gland finally reaches the mouth along a large duct. This arrangement, in which the acini are related to one another like the grapes in a bunch, is called *racemose*. Because the saliva reaches the surface by a duct it is said to be an external secretion and a salivary gland is a gland of *external secretion* or an *exocrine gland*.

After being fixed and stained the glandular cells of the salivary glands are seen to be of two kinds. The serous cells which are small and granular with well stained nuclei produce a clear watery secretion and since the parotid consists almost entirely of such cells saliva from this gland is in general much less viscous but richer in amylase than that from the other salivary glands. The mucous cells which are larger than the serous but clear and transparent are in the majority in the sublingual gland which secretes a thick sticky opalescent material rich in mucin. The submandibular gland has both serous and mucous acini in roughly equal proportions. The greater part of the saliva comes from the submandibular and sublingual glands.

The Nerve Supply of the Salivary Glands.—The salivary glands are supplied by both sympathetic and parasympathetic fibres (Fig. 15, 1). The central origin of the latter, the superior and inferior salivary nuclei, is in the medulla, between the VIIth nucleus and the lateral vestibular (Deiter's) nucleus. The parotid gland receives sympathetic fibres from the superior cervical ganglion by way of the arteries which supply the gland. The parasympathetic fibres pass along the glossopharyngeal nerve and its tympanic branch to the lesser superficial petrosal nerve and so to the otic ganglion. Postganglionic fibres are then distributed through the auriculo-temporal branch of the mandibular division of the trigeminal nerve to the cells of the parotid gland.

The submandibular and sublingual glands are supplied by sympathetic fibres which reach them along the blood vessels from the superior cervical ganglion; their parasympathetic fibres leave the brain in the facial nerve and by its chorda tympani branch, which joins the lingual nerve, and terminate in the submandibular (sublingual) ganglion. From here postganglionic fibres pass to the secreting cells of the submandibular and sublingual glands. Some of the parasympathetic nerve fibres to the submandibular gland, however, pass through the ganglion to synapse with nerve cells in the gland itself. Postganglionic fibres then pass to the secreting cells.

If the chorda tympani (parasympathetic) of a cat or dog is stimulated electrically a copious secretion of saliva occurs at once and may be maintained for a long period. Owing to the marked vasodilatation, the blood flow through the submandibular gland is increased up to fivefold and the gland swells and reddens. The vasodilatation may be so great that the arterial pulsations are transmitted to the veins leaving the gland. The secretion obtained is abundant and watery, with a low content of total solids (1 to 2 per cent.).

After a dose of atropine is given intravenously, stimulation of the chorda produces no secretion but still produces an increase in the metabolic activity of the gland and a marked but smaller vasodilatation. Since various tests show that the preganglionic and postganglionic fibres of the gland are both cholinergic, the absence of blocking of vasodilatation by atropine has been difficult to interpret. Hilton and Lewis have found recently that when the gland cells are activated, either by chorda stimulation or by acetylcholine, they liberate a stable vasodilator substance which is normally carried away from the gland in the lymph. If the lymphatic channels are blocked while the gland is perfused, the venous effluent possesses, during chorda stimulation, vasodilator properties which are not antagonized by atropine or antihistamine drugs. The vasodilatation produced in the submandibular gland on chorda stimulation is not, therefore, due to the action of vasodilator nerves but more probably to the escape from the gland cells into the interstitial spaces of an enzyme which acts upon proteins in the interstitial fluid to form a vasodilator polypeptide which has been called bradykinin.

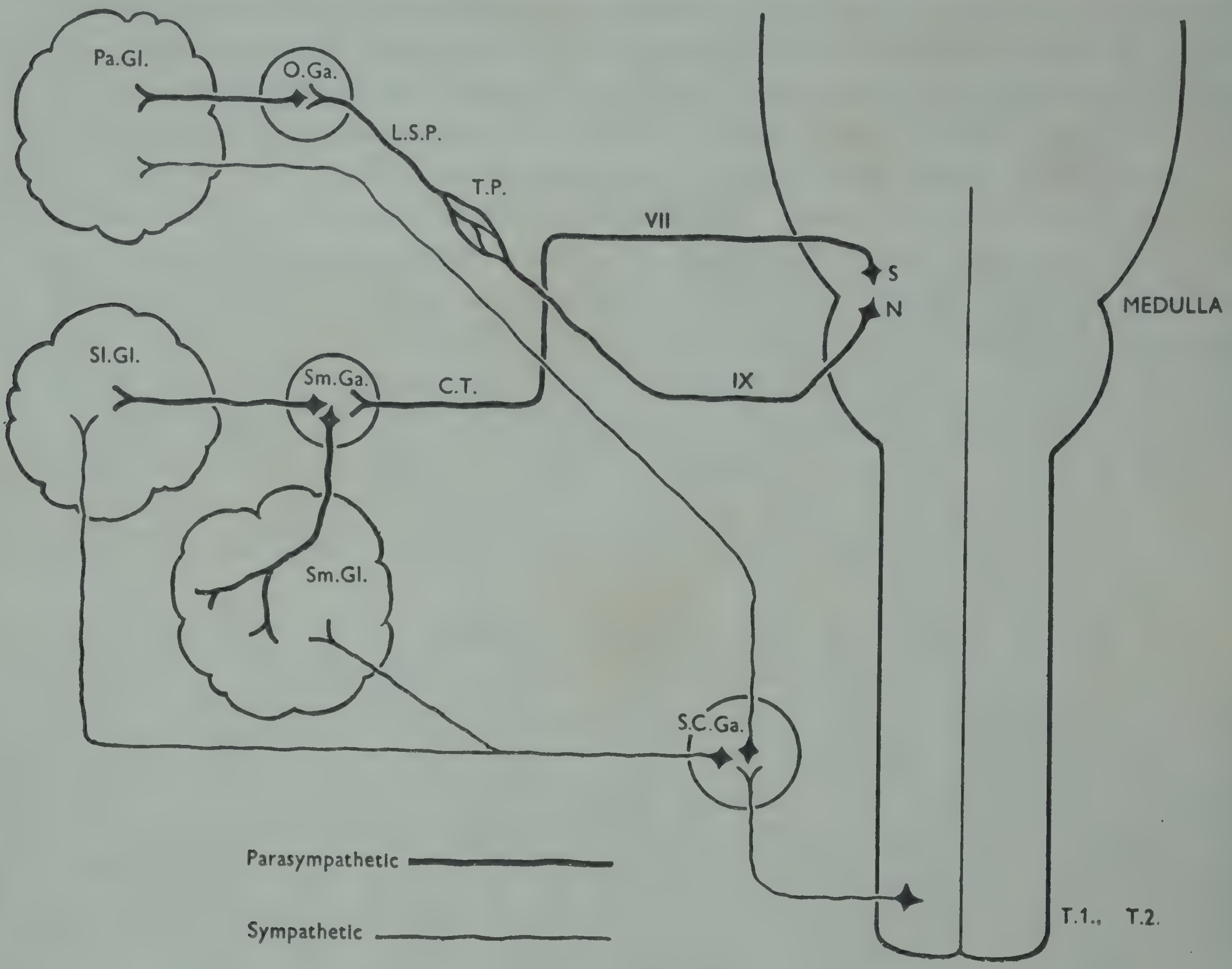


FIG. 15, 1

Diagram to illustrate the nerve supply to the salivary glands. Pa.Gl., parotid gland. Sl.Gl., sublingual gland. Sm.Gl., submandibular gland. O.Ga., otic ganglion. S.C.Ga., superior cervical ganglion. Sm.Ga., submandibular ganglion. L.S.P., lesser superficial petrosal nerve. T.P., tympanic plexus. C.T., chorda tympani. VII, facial nerve. IX, glossopharyngeal nerve. S N, salivary nuclei. T.1, and T.2, first and second thoracic segments of the spinal cord.

Sympathetic stimulation, or injection of adrenaline, produces a small amount of viscous saliva containing a large proportion (6 per cent.) of solids. During stimulation there is vasoconstriction with a reduction in blood flow, but when the stimulation is ended there is a prolonged vaso-

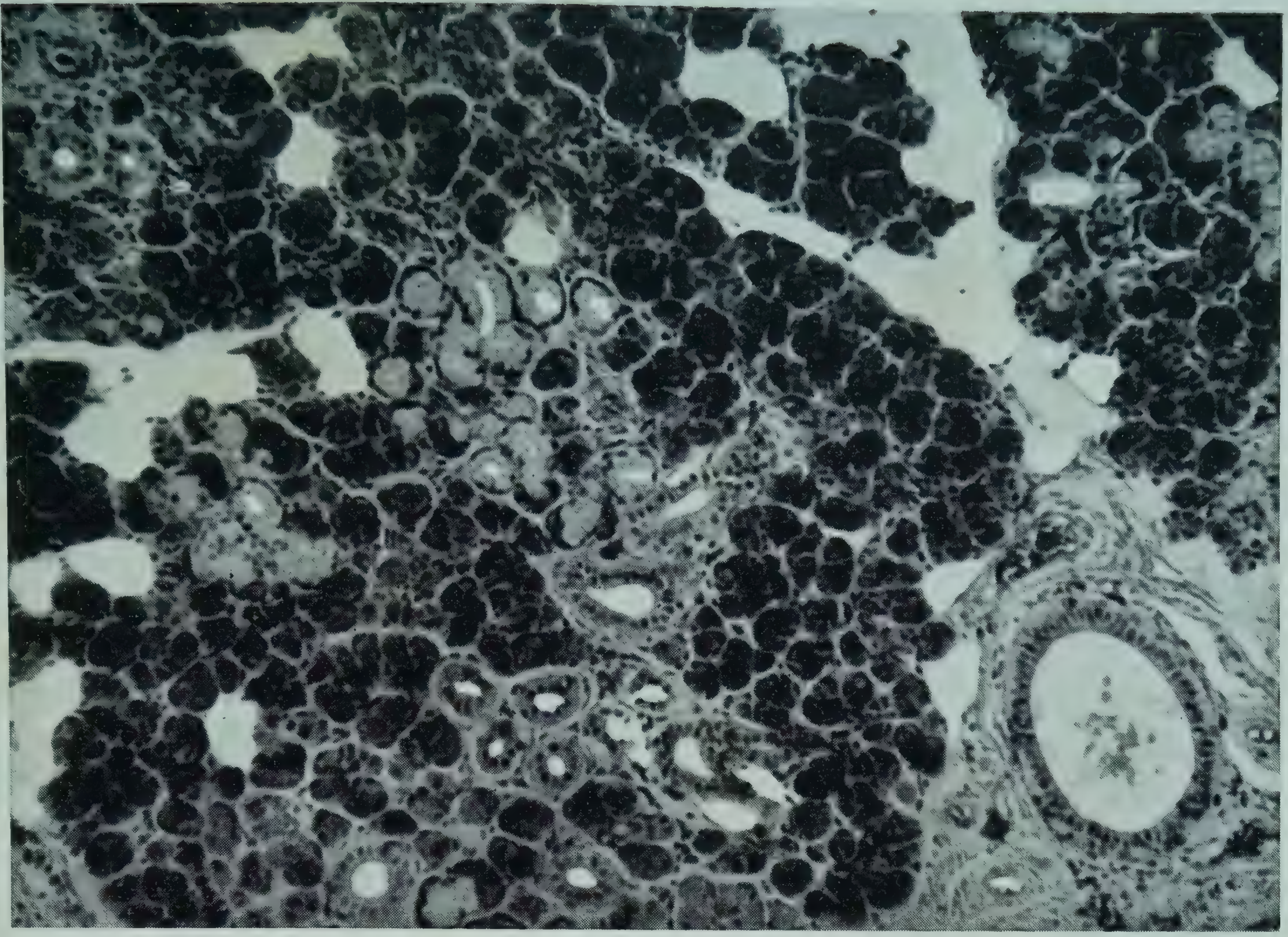


Fig. 15, 2

Human salivary gland (submandibular). The mucous acini are lightly stained. The serous acini are darkly stained. Magnification × 110. Stain: hæmalum, phloxin and tartrazine.

(*By courtesy of A. C. Lendrum.*)

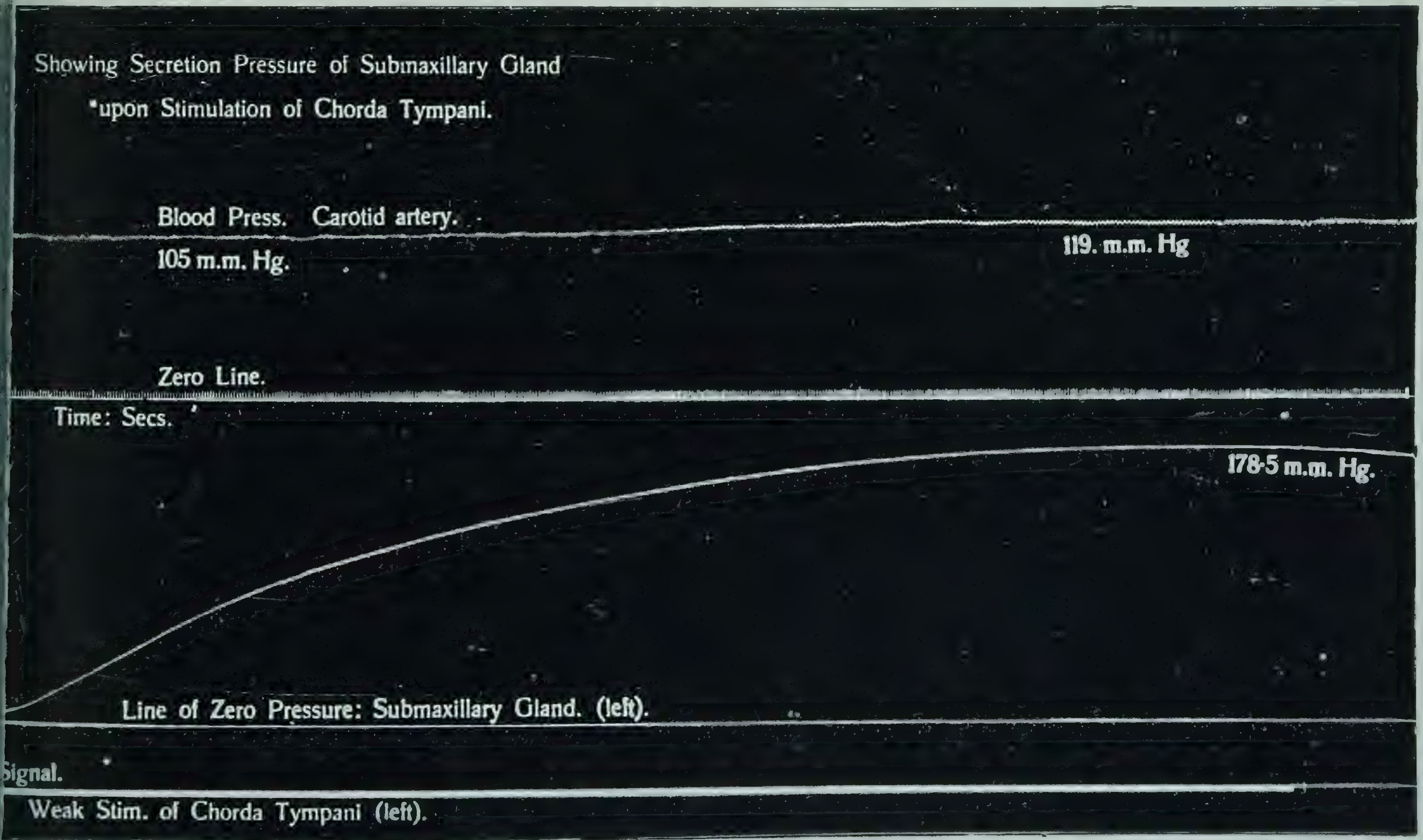


Fig. 15, 3

ecord of the rise of pressure in the submandibular gland during prolonged stimulation of the chorda pani nerve. The submandibular duct was connected directly to the manometer so that no saliva could escape.

(*By courtesy of E. W. H. Cruickshank.*)

dilatation which is paralleled by an increase in the oxygen consumption of the gland.

In the dog, stimulation of the parasympathetic fibres to the parotid gland causes a copious secretion of dilute saliva, whereas stimulation of the sympathetic is followed by little, if any, secretion. If, however, the parasympathetic supply is stimulated immediately after the sympathetic, the saliva produced at first is concentrated and corresponds to that produced from the sublingual and submandibular glands after sympathetic stimulation. This experiment shows that previous sympathetic stimulation, although not producing saliva, has had a pronounced effect on the gland.

Claude Bernard found that when the chorda tympani nerve is cut in the dog a slow secretion of saliva from the submandibular gland begins after a latent period of two or three days, and may continue for five or six weeks until the gland atrophies. This phenomenon is known as *paralytic secretion* and was attributed by Langley to the removal of some inhibitory influence exercised upon the secreting cells through the chorda tympani nerve. It has now been found that paralytic secretion is not continuous but occurs only when the experimental conditions, for example morphine anæsthesia, produce a flow of sympathetic nerve impulses to the gland or a release of adrenaline or noradrenaline from the suprarenals. After section of the chorda the submandibular gland is more sensitive to these substances.

Although there are many drugs which influence both the type and amount of saliva produced, atropine is the only one which is much used in medicine. Atropine, because it blocks the action of acetylcholine, which is the humoral substance of the parasympathetic nerves (Chapter 43), inhibits secretion. Acetylcholine, pilocarpine and eserine all produce a secretion of saliva containing mucus, while adrenaline and ephedrine are said to cause the secretion of thin watery juice.

The Nervous Control of Secretion.—The salivary glands produce only a minimal amount of saliva until food is put in the mouth. The kind of saliva secreted, as well as its quantity, is related to the type of food which is exciting secretion. Thus acid substances produce a rather large volume of watery saliva similar to that excited by dry food whereas milk, or even cold water, produces a much smaller quantity of viscid juice. The enzyme content also varies with the stimulus; thus although meat and weak hydrochloric acid call forth an equal volume of saliva, the enzyme content in the former case is much greater. How these adjustments of salivary juice are made is not known.

The foregoing are all simple reflexes; salivation can, however, be readily conditioned (Chapter 50). Thus the sight, smell, or even the thought of food may produce a copious secretion, sometimes referred to as 'psychic secretion.'

Salivary secretion is reduced, if not actually suppressed, during states of tissue dehydration when for some reason the body has lost large quantities of fluid. The tongue and mouth become dry and a sensation of thirst is produced. The secretion of saliva is also inhibited by emotional stress.

Mechanical stimulation of the œsophagus produced, for example, by the passage of a rubber tube, and irritation of the gastric and duodenal mucous membranes all call forth reflexly a secretion of saliva. Excessive salivation may thus be a symptom of disease in the upper alimentary tract.

Composition of Saliva.—The mixed saliva from the three glands is a viscous fluid which is colourless and opalescent, with a specific gravity of 1·003 to 1·010 and a slightly acid reaction (pH 6·3 to 6·8). The parotid saliva is more acid (average pH 5·8) than the saliva produced by the other two glands. Under the microscope the fluid shows nucleated squamous cells from the buccal lining, together with occasional leucocytes and gland cells, as well as numerous micro-organisms of all descriptions. The total amount produced per day is quite large, about 1 to 1·5 litres in man, with 0·5 per cent. of total solids. The chief organic constituents are mucin and the enzyme ptyalin (salivary amylase), together with small amounts of albumins and globulins, urea, and a trace of potassium thiocyanate.

The mucin of the saliva is a glycoprotein which confers viscosity and lubricating properties on the saliva. Ptyalin which catalyses the breakdown of starch to maltose is activated by chlorides and its optimum pH is 6·7. The organic constituents account for some 60 per cent. of the total solids; the remaining 40 per cent. includes such inorganic substances as potassium, sodium, calcium, chlorides, and especially phosphates and bicarbonates. If there is a rise in the pH of the fluid in the mouth, as can easily happen by loss of carbon dioxide, then calcium carbonate and calcium phosphate, which are soluble in acid but insoluble in alkaline media, are deposited on the teeth as tartar. A rise in pH of the salivary secretion, produced for example by bacterial infection, may lead to precipitation within the duct of calcium salts, with the formation of a salivary calculus.

Functions of Saliva.—The saliva subserves a number of functions. (1) It moistens dry food and facilitates swallowing. Since water evaporates slowly from saliva it prevents desiccation of the oral mucosa. (2) It provides an enzyme for the digestion of starch. (3) It keeps the mouth and teeth clean. During fever, when salivary secretion is suppressed, the lips, teeth and mouth become coated with a mixture of food particles, dried mucus and dead epithelium which, if it is not removed mechanically, speedily becomes the site of bacterial infection. (4) Saliva acts as a vehicle for the excretion of certain substances, for example lead, mercury and iodides. (5) By facilitating movements of the tongue and lips it makes rapid articulation possible. Extreme nervousness, or an injection of atropine, may inhibit secretion to such a point that rapid speech becomes impossible. (6) Saliva subserves the sense of taste by acting as a solvent. The taste buds can be stimulated only when the sapid substance is actually in solution (Chapter 38).

Salivary Digestion.—The motions of chewing have the effect of thoroughly mixing the food with the salivary secretion. In this way ptyalin, the only digestive enzyme contained in saliva, is brought into intimate contact with the starch of the food. The food normally remains for too short a time in the mouth to allow the digestion of starch to proceed very far, but even after the food is swallowed the pH remains for some time favourable to the continued action of amylase and the starch molecules are hydrolysed to dextrins and then to maltose. Hydrolysis is finally arrested by a fall in pH as the acid gastric juice gradually penetrates through the food mass in the body of the stomach.

Mechanism of Secretion.—The fact that a very large and continuous ecretion of saliva can be obtained by stimulation of the parasympathetic erves shows that there must be a considerable transference of fluid Fig. 15, 4) from the blood in the capillaries around the secreting acini hrough the capillary endothelium into the tissue fluid, through the basement membrane to the secreting cells, into the acini and onwards into he ducts. As already explained, alterations of blood flow cannot account or the changes in production of saliva. During stimulation of the chorda tympani a manometer placed in the duct of the submandibular gland may show a secretion pressure greater than the systolic pressure in the carotid artery (Fig. 15, 3). Since the pressure in the capillaries around the acini must be much lower than this, the secretion of saliva cannot be explained as a simple process of filtration from the blood plasma. The saliva is a very dilute fluid with about half the osmotic pressure of blood plasma. If saliva is in fact formed from blood by passage of fluid and salts through a semipermeable membrane, then a very considerable amount of osmotic work must be done by the secreting cells to produce this more dilute fluid. Only the first stage of the process, that is the passage of fluid through the endothelial wall of the capillaries into the tissue spaces, is to be explained on simple physical principles (see Tissue Fluid, Chap. 35).

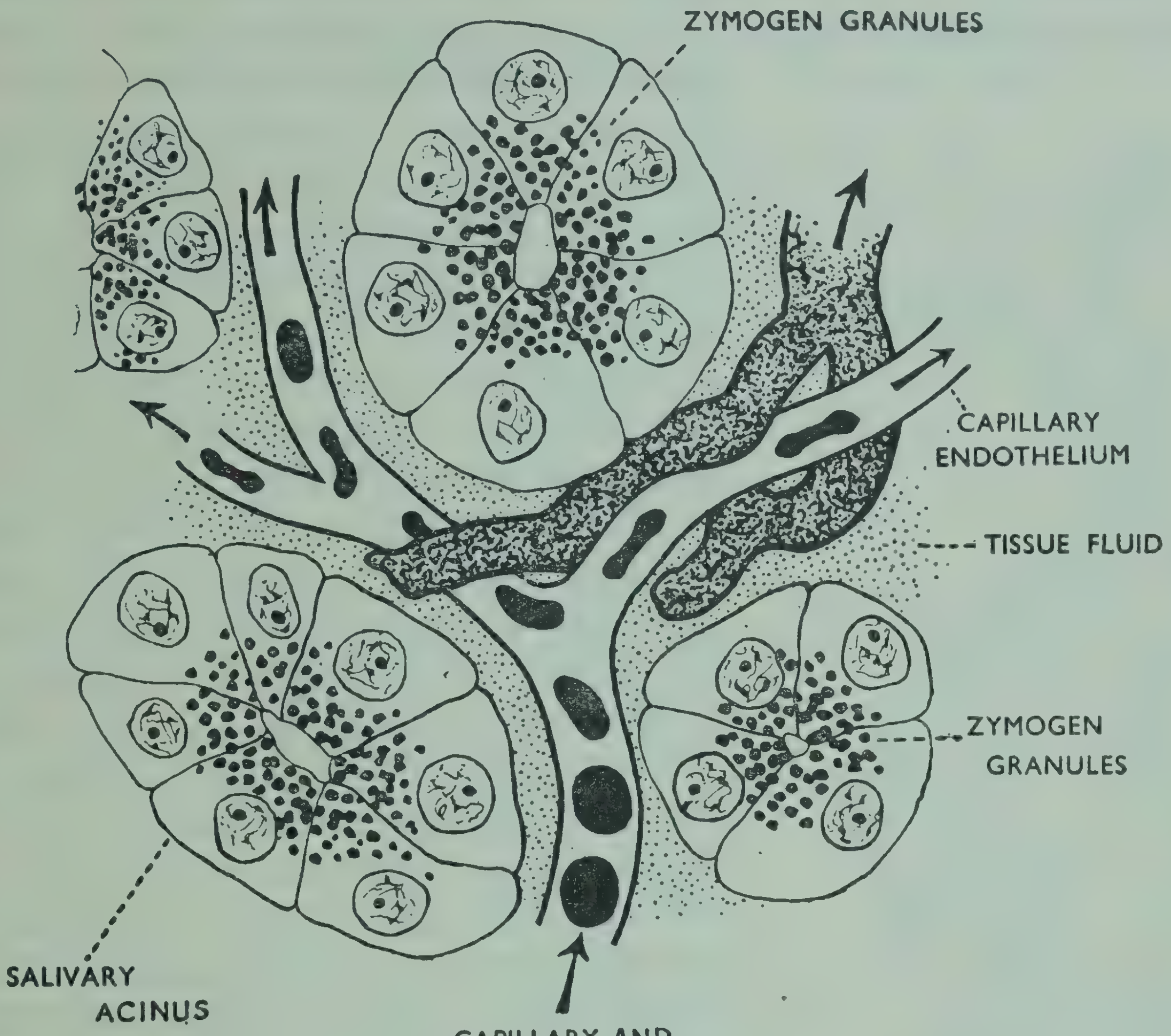


FIG. 15, 4

This diagram shows the relation between the blood vessels and secreting cells of a salivary gland. The blood vessels bring fluid which passes through the capillary endothelium to the tissue fluid. From there it passes into the periphery of the cells of the acinus. Excess of tissue fluid is drained off in the lymphatics (dark stipple). (*After* Cowdry.)

If the volume of the submandibular gland is recorded during stimulation of the chorda tympani it is found that there is an initial shrinkage of the gland in spite of the increased volume of its blood vessels. Thus the first effect produced by stimulation is a discharge of the fluid contained in the acinar cells (thus accounting for the very short latent period); this fluid is then made good from the tissue fluid. This is in marked contrast to the findings after atropinization, when stimulation of the chorda tympani causes an increase in volume of the gland due to vasodilatation and probably to an increase in tissue fluid.

Histological Basis of Secretion.—When living cells from a salivary gland are teased out in blood serum it will be found that the cells are filled with granules, relatively large in the mucous glands and finer in the serous glands. After a period of secretion the cells are found to be smaller and they contain fewer granules, which tend to be collected near the free margin of the cell. Secreting cells in various parts of the body contain granules which are the specific secretion of the cell in an inactive form. If the specific secretion is an enzyme the granules are called zymogen granules (Fig. 15, 5).

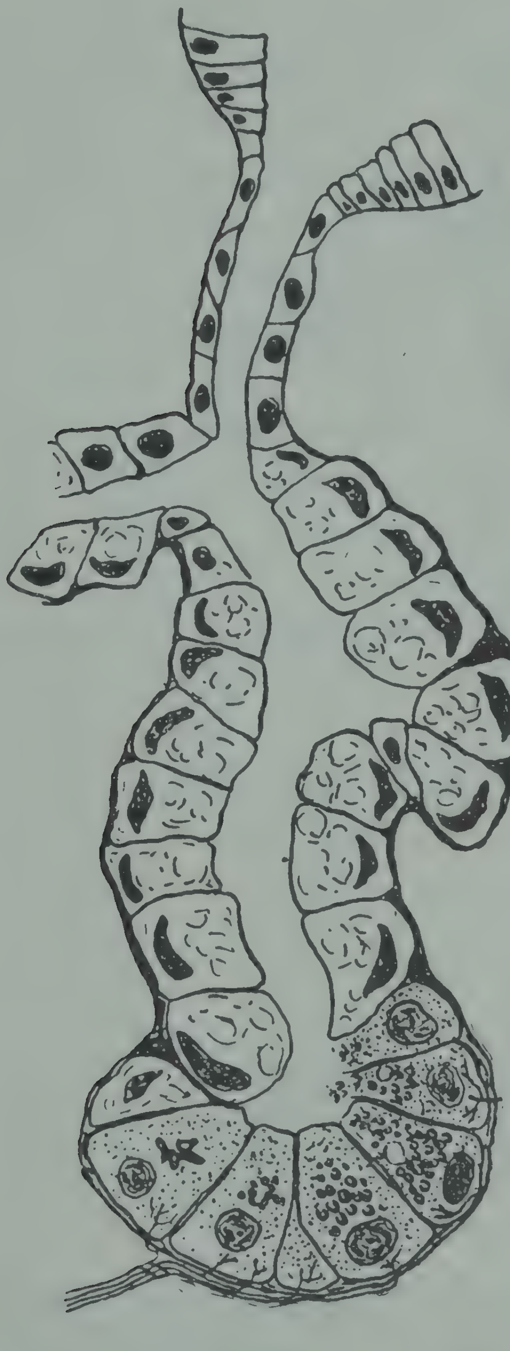


FIG. 15, 5

Section of a submandibular gland to illustrate the method of production of saliva by the acinar cells. The series of intracellular changes described on this page are illustrated in the lowermost cells from left to right. The granules become smaller and more numerous and are finally discharged.

The granules in the serous cells of the salivary glands presumably contain a precursor of ptyalin which is discharged into the acinus as zymogen and later converted to the active enzyme. In the case of the mucous cells there is no doubt that the granules contain some precursor of mucin, called for convenience *mucinogen*, and not mucin itself. Mucin is coagulated by acetic acid whereas mucinogen is made to swell by acetic acid, or any of the usual histological fixatives, thus giving the characteristic clear appearance seen in microscopic preparations.

In the process of secretion the first change is in the granules which absorb water from the surrounding cytoplasm and then go into solution. When this solution is discharged from the cell into the lumen of the gland the remaining cytoplasm is left relatively concentrated, with an osmotic pressure higher than that of its surroundings, and water, therefore, moves in from the tissue fluid. Thus the osmotic pressure of the tissue fluid is raised, and to compensate for this change water passes into the tissue fluid from the capillaries.

Since energy for these changes is obtained by the breakdown of large molecules, breakdown products of low molecular weight pass out of the cell and increase further the osmotic pressure of the tissue fluid which may, therefore, accumulate because its volume is in excess of that required for salivary secretion. The excess fluid is taken up by the lymphatics

(Fig. 15, 4). Part at least of the energy for secretion comes from glucose, the utilization of which increases during activity. Large amounts of phosphate are taken up by the gland during secretion, presumably to form phosphate esters.

During rest the granules re-form. The restitution process is probably fairly slow, since it has been shown that under continuous chorda stimulation the output of mucin falls steadily and the mucinogen content of the cells may take several days to rise to normal.

THE STRUCTURE OF THE ŒSOPHAGUS

The œsophagus is a tube about 25 cm long and about 2 cm in diameter joining the pharynx to the cardiac orifice of the stomach (the cardia). The stratified squamous epithelium lining the buccal cavity is continued through the pharynx into the œsophagus, at the lower end of which there is a sudden transition to gastric epithelium. In the undistended state the mucosa is thrown into several fine longitudinal folds which disappear when it is distended. The submucous coat contains mucous glands. Outside this there is a considerable amount of muscular tissue (circular fibres internally and longitudinal externally) which is red in colour and chiefly striated in the upper third, but consists almost entirely of visceral muscle fibres in the lower third; both striated and visceral muscle fibres are found in the middle third. The absence of a serous coat together with a relatively poor blood supply accounts for the fact that the œsophagus often heals with difficulty after surgical operations upon it.

DEGLUTITION

The process of deglutition can be studied satisfactorily in man by observing with the aid of X-ray cinematography the swallowing of fluids or solids containing a radio-opaque substance such as barium sulphate. This complex process occupies only a few seconds and thus it is not surprising that the exact details are still open to doubt. In the first stage the movements are under voluntary control. The later stages are carried out involuntarily although they are controlled by the medulla oblongata. The mouth is closed and the bolus, collected on the upper surface of the tongue, is forced past the anterior pillars of the fauces (oropharyngeal isthmus) by an upward and backward movement of the tongue caused mainly by the contraction of the mylohyoid and styloglossus muscles. At the same time the soft palate rises and approaches the posterior pharyngeal wall which is brought forward to close off the nasopharynx. Respiration is inhibited. The larynx begins to rise as the bolus passes over the back of the tongue. From this moment the process of deglutition is involuntary. Since the pharynx communicates with both œsophagus and trachea, the movements occurring next are designed to allow the bolus to enter the œsophagus while avoiding the air passages. The tongue moves back like a piston towards the posterior pharyngeal wall and forces the bolus back against the epiglottis, which arrests it for a short time and then becomes folded to form a cowl-like hood over the laryngeal orifice. The entrance to the larynx is closed by the sphincteric action of the girdle of muscles surrounding it. The food then passes over the lateral edges of the epiglottis in two streams into the part of the pharynx immediately posterior to the larynx and then on into the œsophagus. At this moment the cricopharyngeus (the lower fibres of the inferior constrictor), which forms a sphincter at the 'mouth' of the œsophagus, relaxes. When the bolus is safely past the cricopharyngeus, the larynx drops to its original position, the vocal folds open and the epiglottis quickly resumes its initial position. The cricopharyngeus muscle then closes and prevents air from entering the œsophagus during inspiration. If a number of swallowing movements are made in rapid succession the soft palate and epiglottis do not return to their resting positions until the series is completed.

During the first stage of swallowing the lips and jaws are normally closed. If the lips are not closed swallowing becomes difficult; but if the jaws cannot be closed swallowing is nearly impossible. During dental operations, when the jaws must be kept open, saliva accumulates in the mouth and has to be removed by suction. If a record of respiratory movements is made by means of a stethograph it will be found that breathing is temporarily arrested, no matter what the actual phase of respiration is, during the act of swallowing. During this respiratory inhibition the larynx is drawn upwards—without this movement swallowing is very difficult or even impossible.

The passage of fluids down the œsophagus is remarkably fast and even semi-solid material reaches the lower end of the œsophagus in a few seconds. When fluid is swallowed two sounds may be heard by listening with a stethoscope over the left side of the ninth thoracic vertebra, an area which corresponds to the position of the cardiac sphincter. The first sound occurs when the fluid is pushed through the fauces on to the posterior pharyngeal wall, the second follows five to ten seconds later and lasts a few seconds. If milk mixed with barium sulphate is swallowed, it passes rapidly down to the cardiac sphincter, which it reaches in three or four seconds. During the next two or three seconds it runs slowly through the cardiac orifice into the stomach, so producing the second sound. When the bolus is more solid in consistency, for example like porridge, the speed of descent is only slightly decreased, the slowing up occurring in the lower half of the œsophagus. Although a very dry bolus may take several minutes to reach the stomach, such a delay is rather unusual, since saliva swallowed during the descent of the bolus facilitates its progress. Salivation is produced reflexly by the presence of the bolus in the lower half of the œsophagus; the afferent path of this œsophago-salivary reflex is in the vagus nerve. If the subject is in the horizontal position the bolus passes somewhat more slowly than when the subject is standing. When the subject is inverted (head down) the bolus can be seen by X-rays steadily ascending the œsophagus at a rate about one-third of that in the normal vertical position, but during this ascent of the œsophagus peristaltic movements have not been described.

The waves of contraction which carry the food along the œsophagus have recently been investigated by passing fine gum-elastic tubes into the œsophagus and connecting them to recording manometers. Each propulsive wave passes down about 4 cm/sec, the maximum pressure being between 20 and 50 mm Hg (Fig. 15, 6). A swallowing movement start a wave off down the œsophagus, but a second swallowing movemen occurring within 5 sec does not do so. A similar propulsive wave can b started by injecting air or fluid into the œsophagus. The waves cannot properly be described as peristaltic waves, since they are not passed on from segment to segment like a true peristaltic wave in the intestine. A further difference is that the œsophageal waves are not preceded by a fall of pressure. The contractions are reflexly controlled by sensory impulses travelling in the fifth, ninth and tenth cranial nerves to the medulla which controls the onward passage of the wave of contraction. Œsophageal contractions in the dog are abolished when the vagi, which convey the motor impulses, are cut. The observation that in a man whose pharyngeal mucosa was destroyed no contraction wave was seen suggests that contact of food with the upper pharyngeal wall sets off the contraction wave.

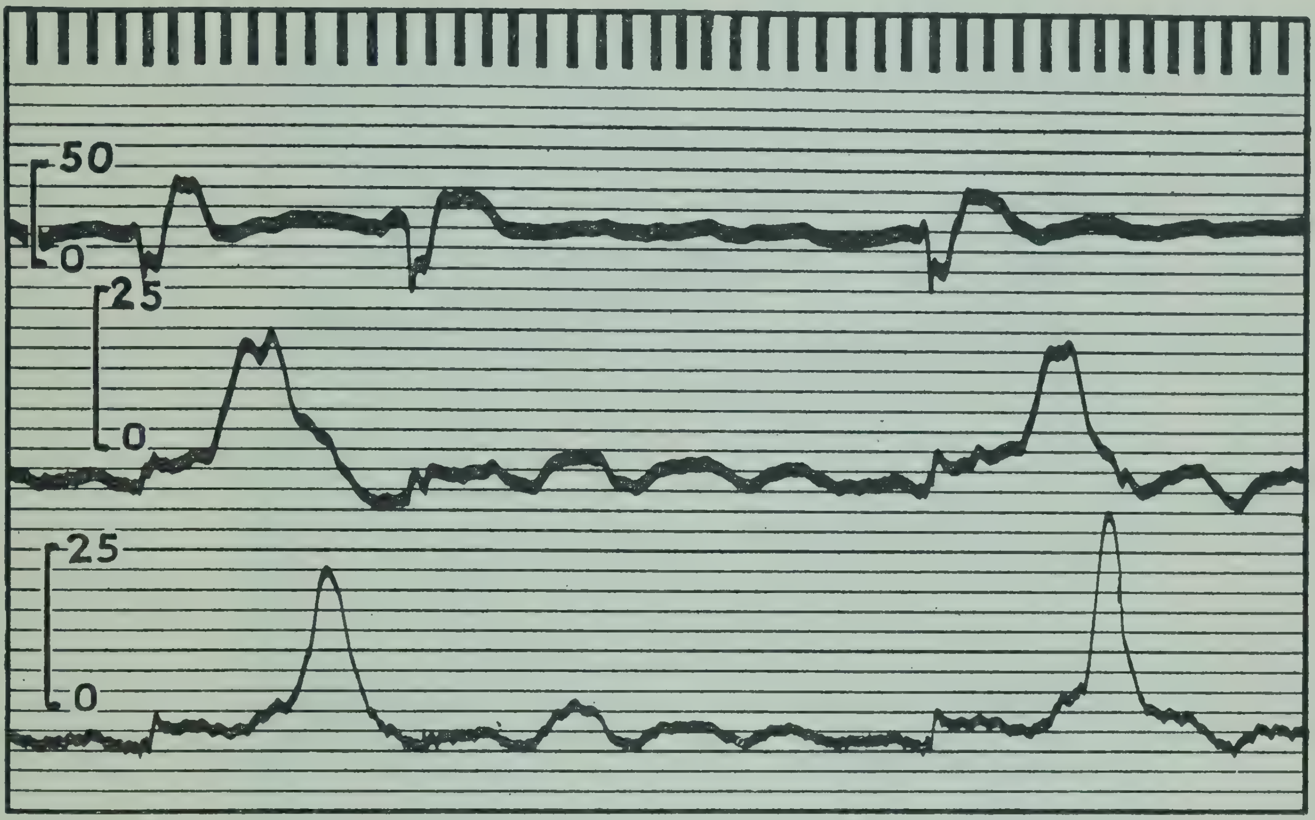


FIG. 15, 6

Pharyngeal swallows and œsophageal propulsive waves. Upper trace: pressure in pharynx. Middle and lower traces: pressures 9 and 18 cm down œsophagus. Note that a pharyngeal swallow repeated too soon fails to initiate a propulsive wave. In this and in Fig. 15, 7, time is marked at the top in seconds and pressures in mm Hg (Dornhorst, A. C., Harrison, K. & Pierce, J. W. (1954). *Lancet*, i, 696).

A small amount of air is swallowed with each bolus and forms the air bell in the fundus of the stomach. Belching is the expulsion from the œsophagus or stomach of this swallowed air rather than of gases produced by putrefactive activity in the stomach. Each eructation of air is followed by swallowing which results in the intake of more air; in certain hysterical states this process may be frequently repeated (*aerophagy*) and lead to actual distension of the abdomen.

When food passes down the œsophagus the cardia gives way when the œsophageal pressure is, say, 4 mm Hg above that in the stomach (Fig. 15, 7). The cardia remains shut, however, when the pressure in the stomach is very much higher than in the œsophagus. On inspiration the pressure in the thorax (p. 540) and, therefore, in the œsophagus falls; at the same time the pressure in the abdomen rises. During a deep inspiration the abdominal pressure may be 15 mm Hg higher than that in the œsophagus, and if an inspiratory effort is made with the glottis closed, the pressure difference may be as high as 80 mm Hg of water without any regurgitation taking place. The cardia thus acts as a valve and not like a true sphincter and in fact dissection of the lower end of the œsophagus and the cardia does not reveal any arrangement of the smooth muscle that could constitute a sphincter. Although the œsophagus passes through the right crus of the diaphragm this muscle seems to play no part in the control of the opening into the stomach; even during vigorous respiratory movements the diaphragm does not press on a finger placed in the œsophagus at operation. There is, however, as explained above, a quite efficient mechanism for preventing reflux of gastric contents into the lower end of the œsophagus; if this mechanism is defective reflux of acid damages the

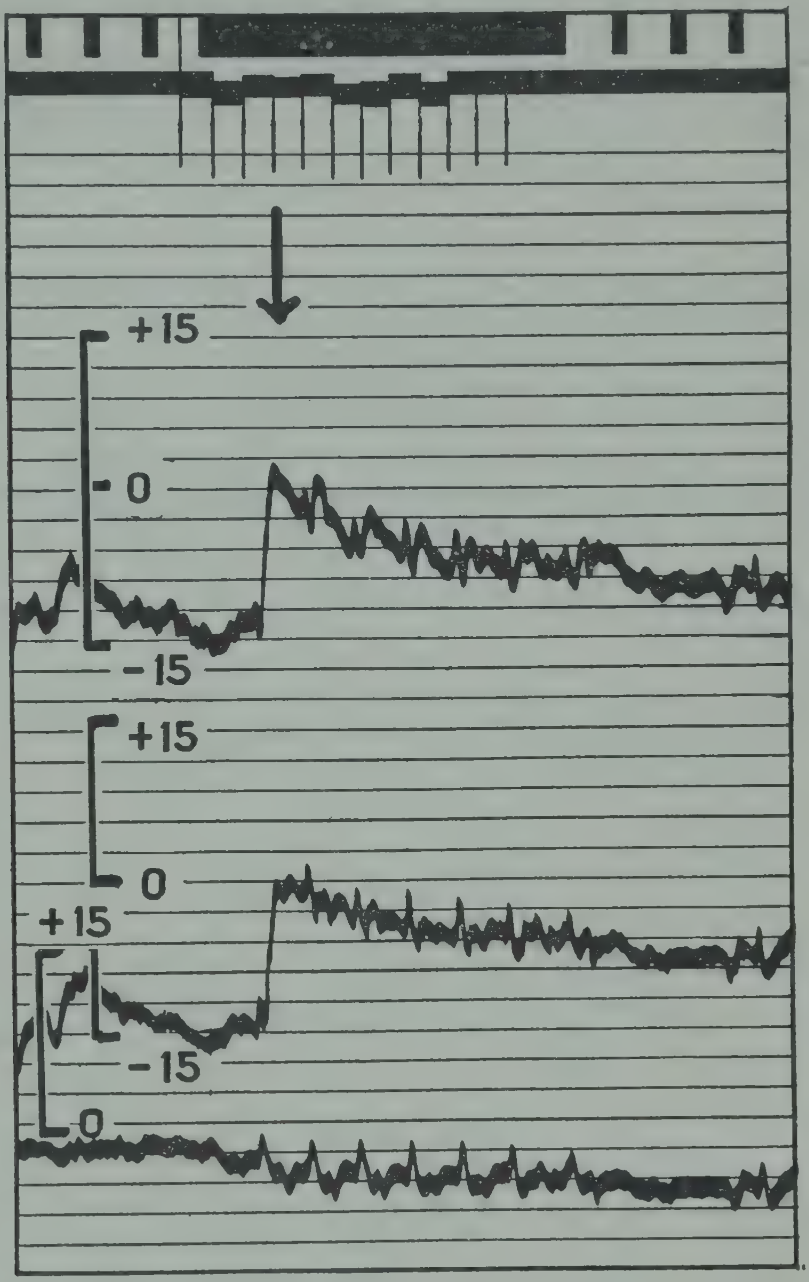


FIG. 15, 7

Upper two traces: pressure in lower œsophagus. Lower trace: pressure in stomach. Barium drunk during long signal shown at top; films exposed at short signals shown below this. Arrow indicates timing of first film showing barium passing into stomach; at the moment the pressure in the stomach is about −5 mm Hg and in the œsophagus nearly zero (that is atmospheric), (Dornhorst, A. C., Harrison, K. & Pierce, J. W. (1954). *Lancet*, i, 698).

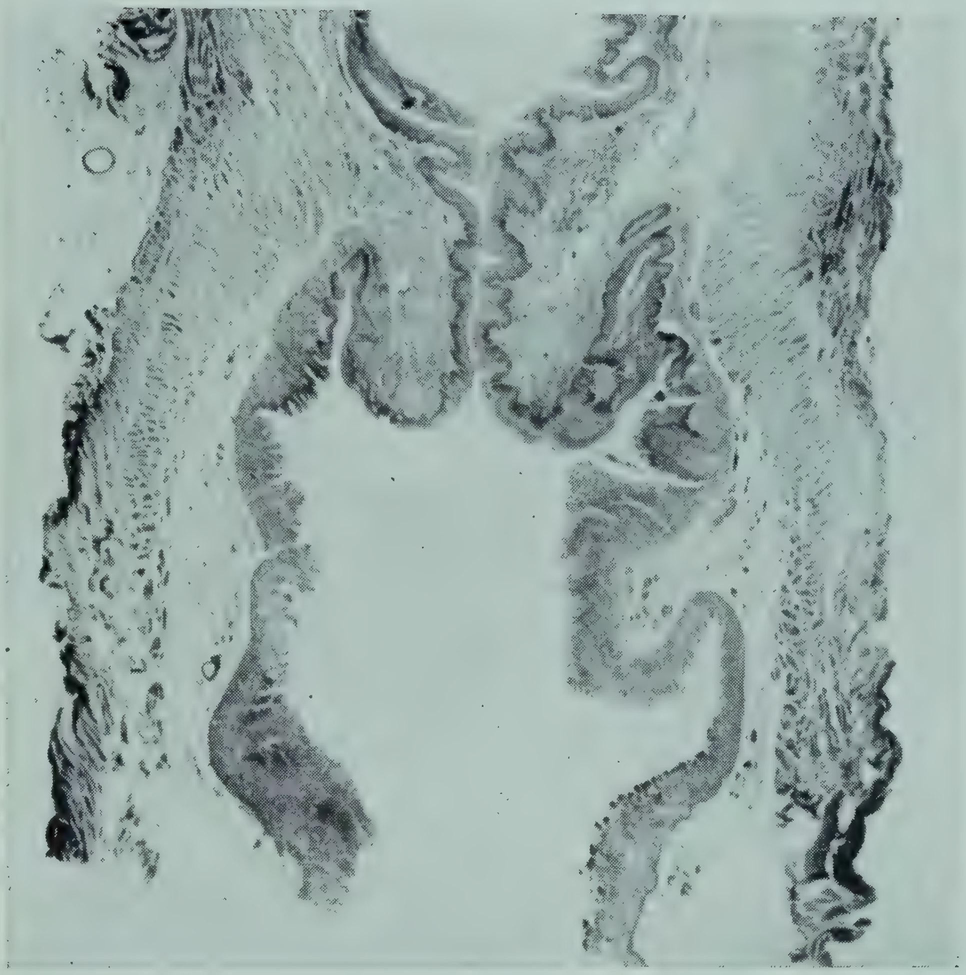


Fig. 15, 8

Section of cardiac area of œsophagus and stomach from operation specimen showing mucosal folds forming a valve; in this case the folds are more prominent than usual (Creamer, B. (1955), *Lancet*, i, 590).

œsophagus and the subject has a burning sensation (heartburn) about 30 seconds after the entry of gastric fluid. It seems likely that this valvular action depends on the structure of the lax mucosa of this region (see Fig. 15, 8); the muscularis mucosæ does not seem to be involved for the following reasons. Irritation of the mucosa by carminatives such as tincture of cardamom causes hyperæmia (vasodilatation) with thickening of the mucosa and impairment of its valvular action. But in cases of mucosal atrophy (for example pernicious anæmia) in which the muscularis mucosæ is intact, carminatives do not interfere with the valvular action presumably because the atrophied mucosa cannot swell. Reflux into the œsophagus is common after a meal and could be attributed to hyperæmia of the mucosa. The valve cannot, however, be a simple mechanical obstruction to flow since it must be capable of active opening when vomiting occurs.

REFERENCES

Ardran, G. M. & Kemp, F. H. (1952). The protection of the laryngeal airway during swallowing. *Brit. J. Radiol.* **25**, 406–416.

Dornhorst, A. C., Harrison, K. & Pierce, J. W. (1954). Observations on the normal œsophagus and cardia. *Lancet,* i, 695–698.

Lerche, W. (1950). *The Œsophagus and Pharynx in Action.* Springfield, Illinois: Thomas.

Negus, V. E. (1950). Pharyngeal diverticula. Observations on their evolution and treatment. *Brit. J. Surg.* **38**, 129–146.

O'Rourke, J. T. & Miner, L. M. S. (1951). *Oral Physiology.* London: Kimpton.

Rushmer, R. F. & Hendron, J. A. (1951). The act of deglutition: a cinefluorographic study. *J. appl. Physiol.* **3**, 622–630.

CHAPTER 16

THE STOMACH

GENERAL STRUCTURE OF INTESTINAL TRACT

A GENERALIZED plan of the structure of the abdominal part of the intestinal tract is given in Fig. 16, 1. The peritoneum is reflected off the posterior wall of the abdomen as the mesentery, and then on to the gut to form the external serous coat. Immediately underneath this serosa there is a longitudinal layer of smooth muscle with a circular muscle coat internally. The fibres of this coat are not arranged in a truly circular fashion but in a close spiral which passes anticlockwise from the gastric to the anal end of the intestine. The tissue lying between the circular muscle coat and the muscularis mucosæ is referred to as the submucosa. Finally, bordering the lumen of the intestine is the mucous coat, consisting of the surface layer of epithelium with subepithelial connective tissue containing the muscularis mucosæ. The myenteric nerve plexus (Auerbach's plexus) lies between the two muscle coats and the nerve plexus of the submucosa (Meissner's plexus) lies, as its name indicates, in the submucosa.

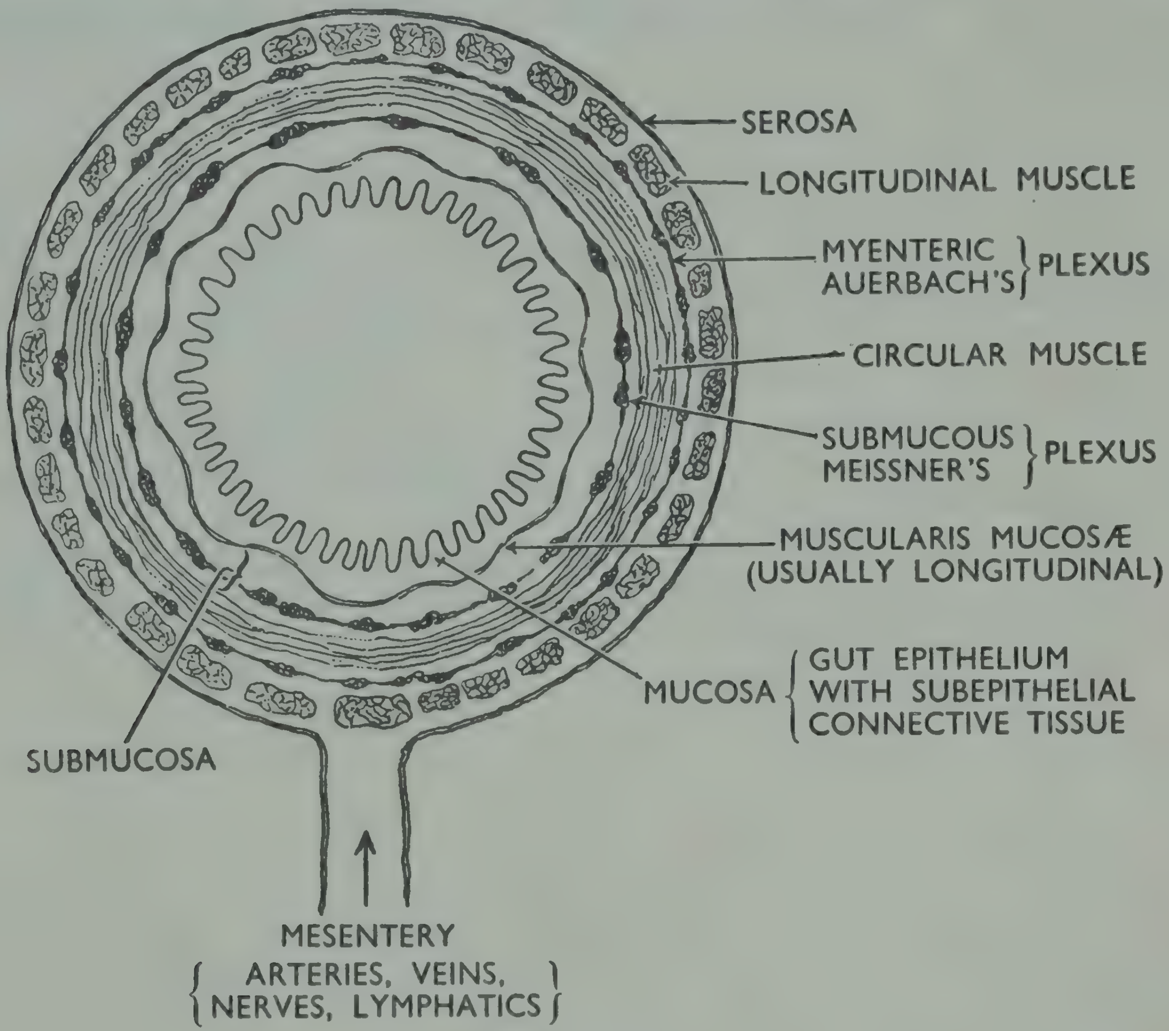

SEROSA
LONGITUDINAL MUSCLE
MYENTERIC AUERBACH'S } PLEXUS
CIRCULAR MUSCLE
SUBMUCOUS MEISSNER'S } PLEXUS
MUSCULARIS MUCOSÆ (USUALLY LONGITUDINAL)
MUCOSA { GUT EPITHELIUM WITH SUBEPITHELIAL CONNECTIVE TISSUE
SUBMUCOSA
MESENTERY
{ ARTERIES, VEINS, NERVES, LYMPHATICS }

FIG. 16, 1
Structure of the intestine.

STRUCTURE OF THE HUMAN STOMACH

The muscular coat of the stomach is composed of three layers: (1) an external longitudinal layer (immediately under the serosa) part of which is continuous with the longitudinal fibres of the œsophagus; (2) a circular layer (continuous with the circular layers of the œsophageal muscle) which increases in thickness as it passes towards the pylorus where there is a pronounced local thickening, the so-called pyloric sphincter; (3) in the proximal portion only, a well-developed oblique layer lying inside the circular layer. The mucosa is thicker in the pyloric region than in the fundus. When the stomach is empty the mucosa, being loosely attached to the muscle tissue, falls into folds (*rugæ*) which disappear when the organ is distended. These folds are partly due to contractions of the muscularis mucosæ. If a piece of excised stomach is perfused with Locke's solution containing tiny glass spheres it is found that spheres up to 160 μ in diameter pass out in the venous effluent. Arteriovenous anastomoses up to this size must therefore be present in the submucous coat of the human stomach. The lymphatics of the stomach consist of a fine plexus of freely anastomosing channels and thus lymph flowing in any large lymph vessel passing posteriorly from the stomach to the cysterna chyli may be derived from a large area of the stomach.

GASTRIC MOVEMENTS

The shape and movements of the human stomach are best studied by observing the X-ray appearances after administration of a meal containing

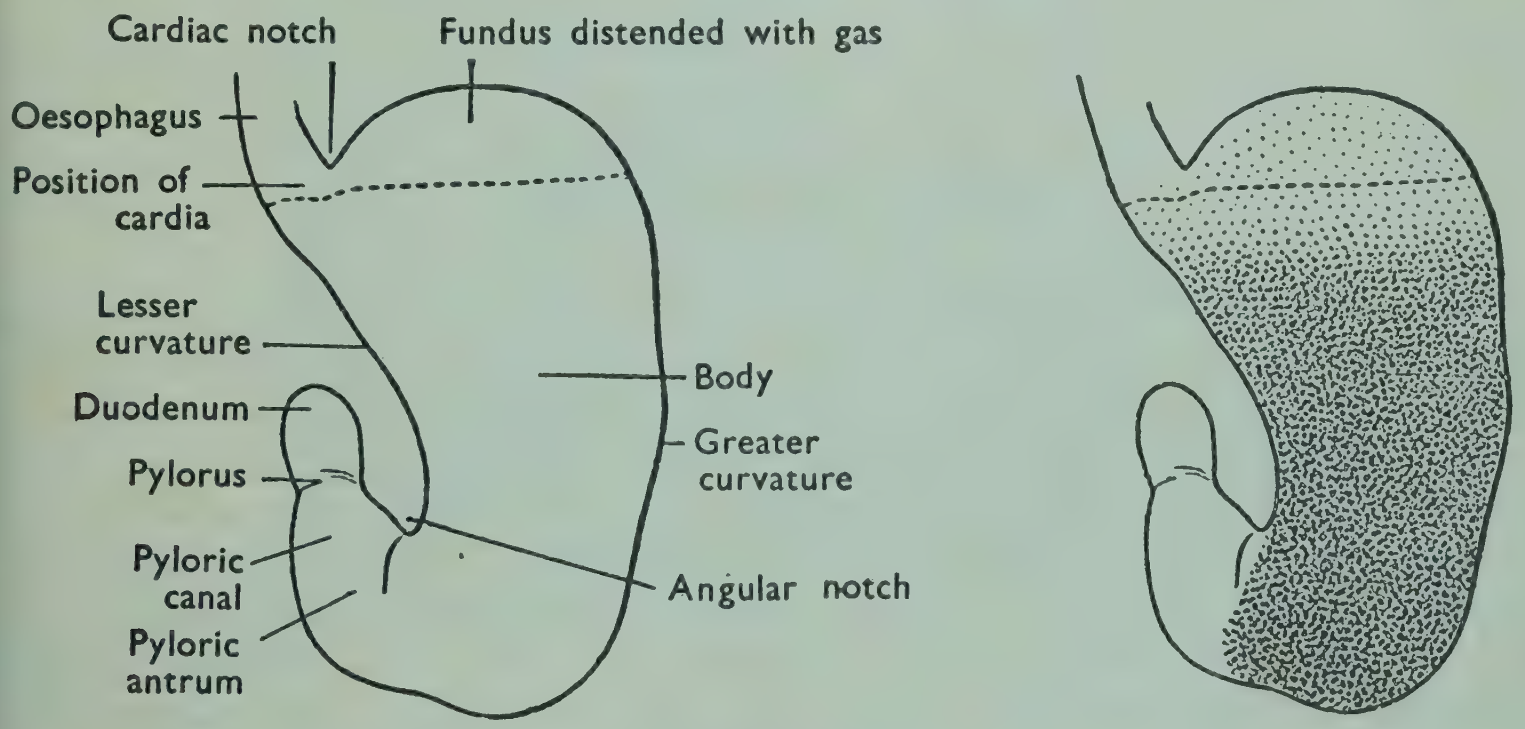


Fig. 16, 2

An outline of the full stomach to show the names of the various parts. The dotted line represents the level to which a large barium meal fills the stomach. The shaded area of the diagram on the right indicates the density of distribution of oxyntic cells according to Berger. There are almost none in the antrum and relatively few in the fundus. (Adapted from *Gray's Anatomy*, 30th ed., 1949, p. 1364. London: Longmans Green.)

barium sulphate, which is opaque to X-rays. This method shows that the stomach has a **J**-shape when the subject is standing (Fig. 16, 4 and 5) and a **U**-shape when he is lying. There is considerable variation in the shape and position of the normal stomach according to the position of the subject, his skeletal build, the tone of the muscular wall and the state of the adjacent viscera.

As food enters the stomach and fills it up the smooth muscle fibres increase in length and allow the stomach walls to relax, so that the intragastric pressure remains nearly constant at a very low value until the stomach is very large (*receptive relaxation*). This maintenance of a slight but constant pressure is referred to as tonus. When the stomach of an animal has been outlined with silver-wire sutures which can be seen in an X-ray photograph it has been found that the lesser curvature remains nearly constant in position and contour, whereas the greater curvature of the body and the fundus increase markedly in length as the organ becomes distended.

When the stomach contains a suspension of barium sulphate peristaltic waves can easily be seen on X-ray examination. They begin in the body of the stomach and travel towards the pylorus as shallow indentations becoming deeper in the pyloric antrum where they may appear to bisect the organ. As a rule from one to three such waves can be seen at one time. They are produced by progressive contractions of the circular muscle fibres of the stomach occurring about three times per minute. The question as to whether similar movements occur in the empty stomach or whether another type of activity, the so-called hunger contractions, exists has not yet been satisfactorily determined. The only method of study that does not require the introduction of barium or a recording balloon into the stomach, that is to say observations of the movement of wire sutures or lead-shot embedded in the wall of the stomach of animals, reveals no such activity.

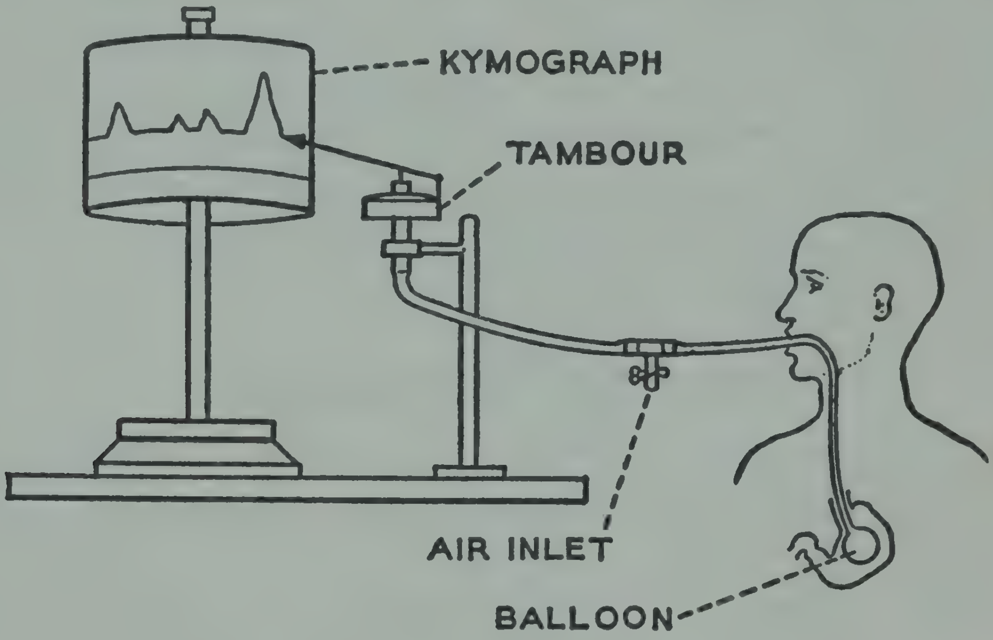


FIG. 16, 3

Method of recording gastric contractions (A. Kay (1947), *Lancet*, i, 448).

[Since vigorous peristalsis rarely occurs in the upper two-thirds of the stomach this part can be regarded as a storage organ in which the food is moistened and softened and mixed with gastric juice which digests the protein. The food is churned up in the pyloric part of the stomach from which small amounts of the contents pass from time to time into the duodenum.]

The activity of the intact human stomach has often been studied by the arrangement described in Fig. 16, 3 but it is very doubtful whether the records produced represent gastric movements at all accurately and they may in fact be highly fallacious. The effects of certain drugs on

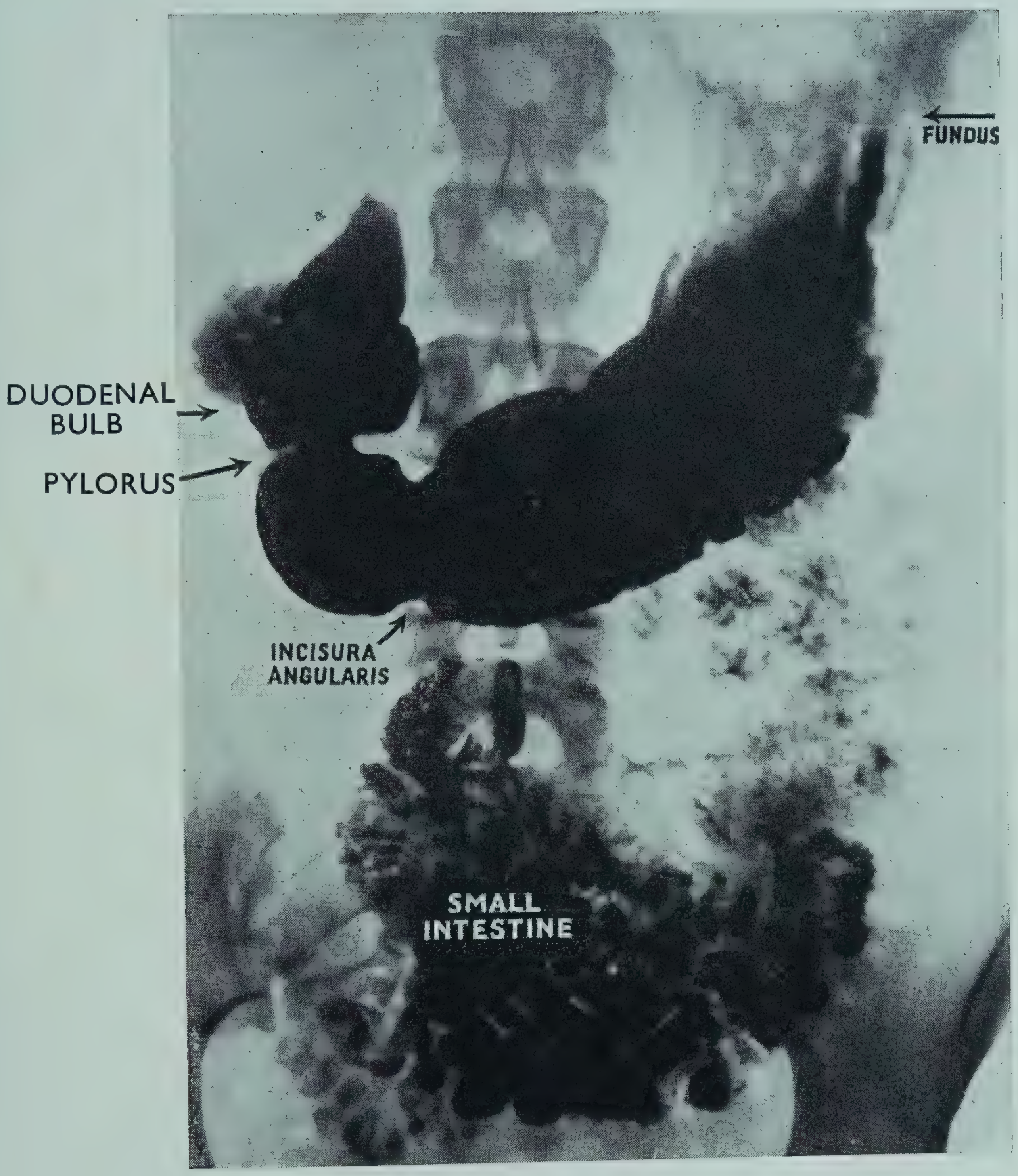

FUNDUS
DUODENAL BULB
PYLORUS
INCISURA ANGULARIS
SMALL INTESTINE

Fig. 16, 4

X-ray photograph of abdomen twenty minutes after a barium meal. Patient erect.

(*By courtesy of R. C. Garry.*)

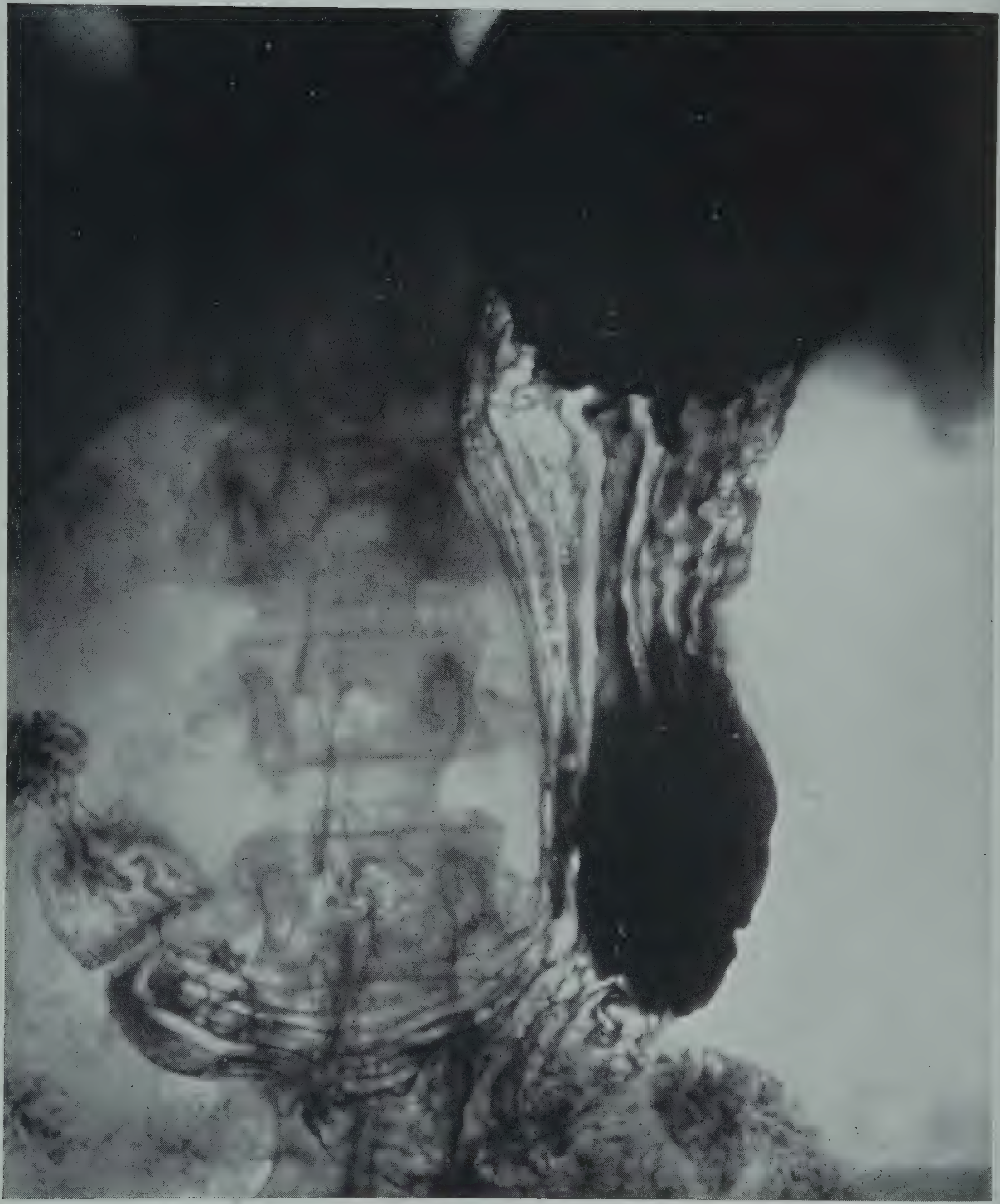


Fig. 16, 5

X-ray photograph of stomach taken twenty minutes after a barium meal (micropaque) to show the mucosal pattern of stomach and duodenal cap. Patient supine.

(*By courtesy of C. Pickard.*)

gastric motility can, however, be demonstrated qualitatively in this way (Fig. 16, 7). Information about gastric activity has been obtained by direct observation in a few persons who, because of a stricture of the œsophagus or of damage to the abdominal wall, have acquired an opening through the anterior abdominal wall into the stomach (*gastric fistula*), and Wolf and Wolff made many observations on their subject, Tom. They found that gastric motility was scarcely affected by mechanical stimulation of the mucosa but it was inhibited by chewing movements, by exposure to cold and by fear, or by weak alkali or strong acid introduced through the fistula. It is now possible to observe and photograph a small area of gastric mucosa through a slightly flexible tube (gastroscope) introduced into the stomach along the œsophagus. This instrument gives little information about gastric movements but it can sometimes reveal the site and nature of disease.

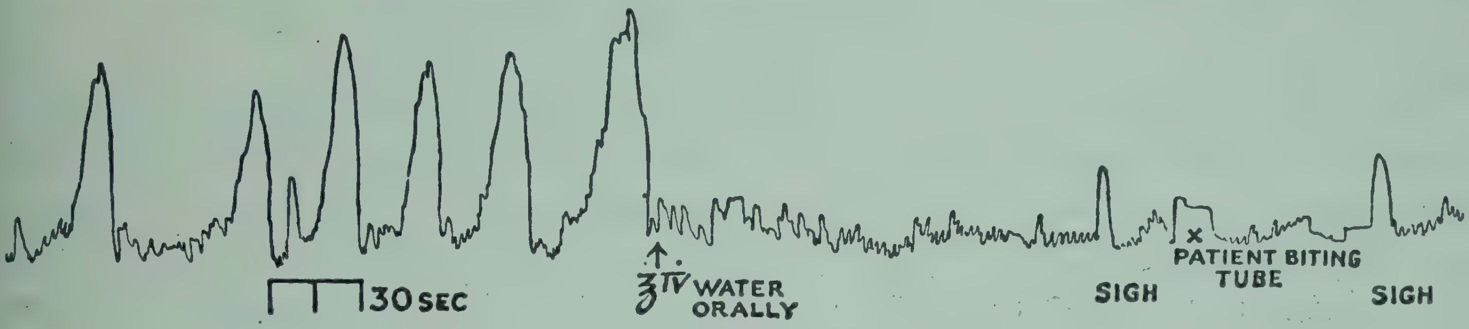


Fig. 16, 6

Record of contractions of human stomach obtained from a balloon in the stomach, as in Fig. 16, 3, attached to a water manometer. Time in 30 seconds. At the arrow the subject swallowed 120 ml. water.

(*By courtesy of W. F. Anderson.*)

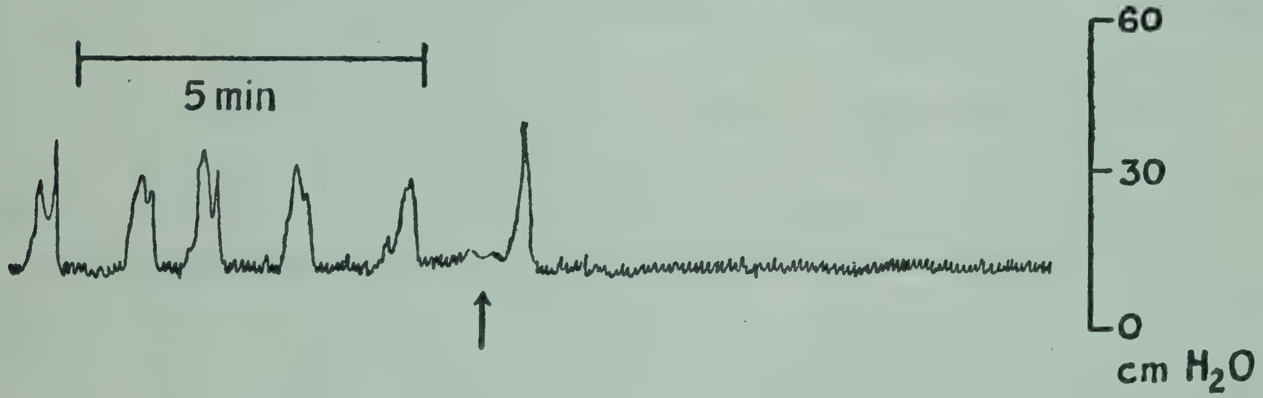


Fig. 16, 7

Record of gastric contractions obtained from a balloon placed in the stomach, as in Fig. 16, 5. 1·5 mg. prostigmine hydrochloride given 35 minutes before the recording began. (Prostigmine is an anticholinesterase.) Effect of 0·5 mg. atropine sulphate (at arrow) on gastric contractions.

(*By courtesy of A. H. James.*)

In dogs, continuous perfusion with acetylcholine or continuous stimulation of the vagus results in hypermotility and hyperacidity, and later to ulceration of the stomach and duodenum. Administration of acetyl-β-methyl choline through the gastric fistula in Tom's case produced no increase in activity, but eserine put directly in the stomach produced very marked and very vigorous contractions which lasted for an hour and a half. These severe contractions were accompanied by cramping pain in the lower left quadrant of the abdomen. Atropine, which blocks vagal effects, arrested gastric contractions for several hours if given in large doses sufficient to cause marked dryness of the mouth. Adrenaline (10 ml. of 1 : 10,000) solution inhibited spontaneous gastric contractions for about a quarter of an hour. In man for some time after double vagotomy gastric motility is decreased and the emptying time is greatly prolonged.

Gastric Emptying Time

The gastric emptying time is the period elapsing between the ingestion of a meal and the time at which the last portion of some easily identified constituent leaves the stomach. The ordinary barium meal used for X-ray examination has completely left the stomach within six hours.

All observers of the activity of the stomach through gastric fistulæ in man have reported that starchy food leaves the stomach quickly; food containing meat passes through relatively slowly. Leafy vegetables produce a further delay and fatty foods pass through at the slowest rate. In a state of fear all gastric functions are delayed and food may remain undigested in the stomach for twelve hours or more; excitement, on the other hand, may reduce the emptying time. Thus both the emotional state of the subject and the composition of the meal have a great influence on emptying time and motility. Animal experiments, described below, have shown that gastric movements are inhibited by the presence of food materials in the small intestine. Since this effect is reduced by section of the vagus nerves the inhibitory effect may be at least partly due to a so-called enterogastric reflex of which the vagus is an essential part.

The 'Pyloric Sphincter' and Control of Peristalsis.—The junction between the stomach and the duodenum is indicated on the external surface by a groove and internally by a change in the appearance of the mucosa. In this region, too, there is some thickening of the circular muscle fibres of the stomach. These observations, as well as radiological findings in the human subject, have suggested that there is in this position a true sphincter that controls the onward passage of food from the stomach into the small intestine. It has recently been found, however, that the pylorus behaves no differently from adjacent muscular structures. Excision of the pylorus does not alter the rate at which the stomach empties and, even if an indwelling cannula is inserted through the pylorus so as to keep it patent, the gastric emptying time is not appreciably altered.

The passage of food, from the stomach into the duodenum, is the result partly of peristaltic waves which begin in the body of the stomach and pass successively over the pyloric canal, the pylorus and the upper part of the duodenum and partly to pressure differences between the stomach and duodenum which are the result of peristalsis. The narrowing of the canal at the pylorus has the effect of reducing regurgitation of the duodenal contents into the stomach (regurgitation does, however, frequently occur) and of preventing the onward passage into the duodenum of large particles of food until they have been broken up by the action of the gastric juice and the churning effect of peristalsis.

There is evidence that peristalsis in the stomach is initiated in a 'pace-maker' in or near the cardia with an autonomous rhythm. Animal experiments have shown, however, that stimulation of the anterior hypothalamus can produce increased peristalsis while stimulation of certain areas of the cortex produces inhibition. [Small peristaltic waves probably occur in the empty stomach though their vigour is increased as the stomach becomes distended with food. The activity of peristalsis is also governed by both humoral and nervous factors. It is now well established that the vigour of gastric peristalsis is reduced by the presence in the duodenum of carbohydrates, fats, the products of protein digestion, hyper- and hypo-tonic solutions and acids. Similar inhibitory effects can be demonstrated in denervated gastric pouches] and in transplanted stomachs and, further, the effect on gastric peristalsis of these substances is abolished by continuous aspiration of the duodenal contents. It is clear, therefore, that these effects can only be brought about by the release of a chemical substance into the circulation. The humoral agent, called enterogastrone, is produced in the wall of the duodenum and upper jejunum and is released into the circulation by many substances, the most potent of which are acid and fat.

We have already seen that in man double vagotomy leads to reduction in gastric motility. In animals distension of the duodenum, as well as mechanical irritation, is followed by inhibition of peristalsis and this is also, in part at any rate, a nervous process (*entero-gastric reflex*) mediated through the vagus nerves. Vagotomy also reduces the responses obtained by the introduction of specific substances into the duodenum.

The onward passage of food from stomach to duodenum is thus a co-ordinated process in which both humoral and nervous mechanisms normally take part.

THE SECRETION OF GASTRIC JUICE

The surface of the mucous membrane of the empty human stomach is thrown into irregular folds (*rugœ*) (Fig. 16, 5) which become flattened out as the organ distends. Close inspection shows that the whole of the surface is uniformly studded with minute dimples which are the openings of innumerable *gastric pits*. The glands of the stomach communicate with the bottom of these pits and discharge their secretion into them. The whole mucous membrane is covered with slimy mucus secreted by the surface epithelium, by the cells lining the gastric pits and by some of the cells in the gastric glands themselves.

The Collection of Gastric Juice.—In early times human gastric juice was collected by the swallowing of sponges which were subsequently withdrawn and squeezed to expel the fluid. The secretion of the gastric mucosa has been observed directly through openings into the stomach. Classical observations of this sort were made by an American physician, Beaumont, upon a French Canadian, Alexis St Martin, who had sustained a gunshot wound of the abdomen which, on healing, left an opening into the stomach. In experimental animals it is possible to make fistulous openings of this sort but a better method was developed by Pavlov who isolated from the fundus a pouch of mucosa which opened upon the abdominal wall. The mucosa of such a Pavlov pouch is isolated from the main part of the stomach but its blood and nerve supplies are intact; it provides an excellent means of collecting gastric juice uncontaminated by

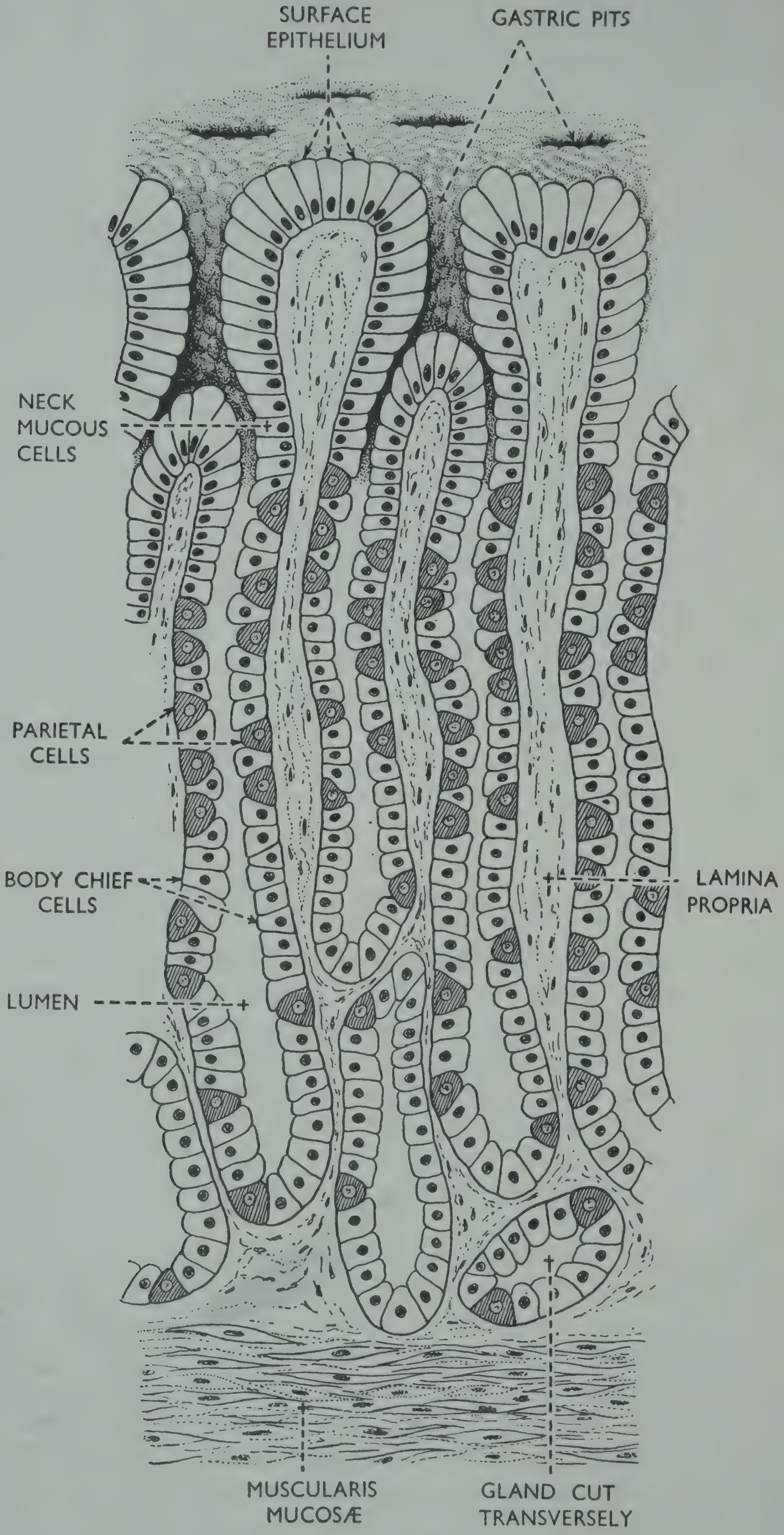

SURFACE EPITHELIUM
GASTRIC PITS
NECK MUCOUS CELLS
PARIETAL CELLS
BODY CHIEF CELLS
LAMINA PROPRIA
LUMEN
MUSCULARIS MUCOSÆ
GLAND CUT TRANSVERSELY

Fig. 16, 8

Diagrammatic section through the human gastric mucosa. Oxyntic cells are shaded. The depth of the mucosa varies between 0·5 mm and 1·5 mm in man.

foodstuffs (Fig. 16, 10). In human subjects gastric secretion can be aspirated through a rubber tube (Ryle's tube) passed from the nose or mouth through the œsophagus into the stomach; however, gastric juice so obtained is mixed with swallowed saliva and any material regurgitated through the pylorus from the duodenum. Secretory activity can also be observed directly through the gastroscope.

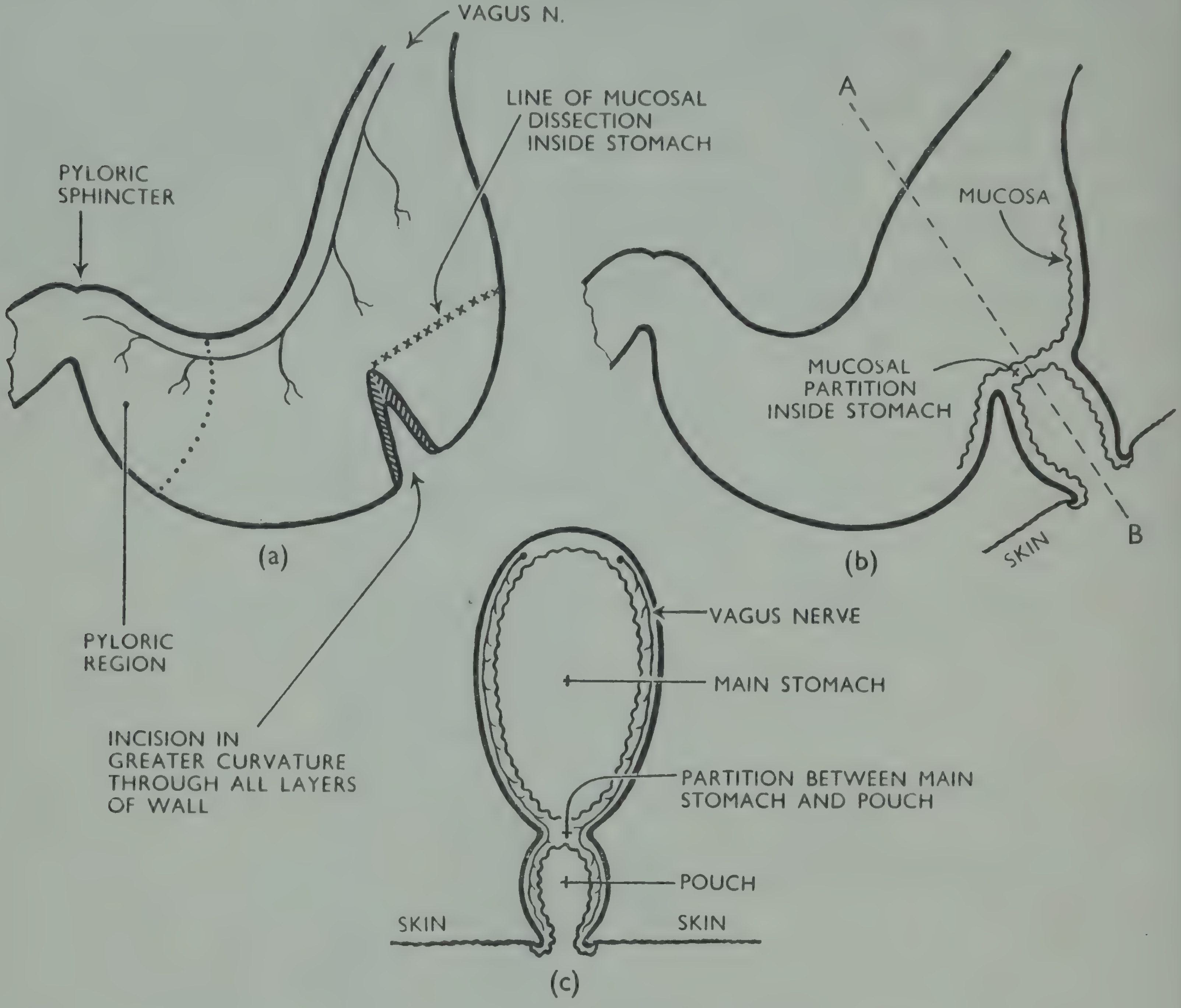
The Composition of Gastric Juice.—Table 16, 9 gives the average composition of gastric juice; some two to three litres are excreted per 24 hours. Gastric juice is usually characterized by a high concentration of hydrochloric acid but it also contains some 0·4 per cent. of organic material, including *mucin* and the enzymes, *pepsin* (and perhaps *rennin*) and *gastric lipase.* The inorganic constituents include sodium, potassium, calcium, magnesium, chlorides, phosphates and sulphates. An important constituent about which, however, relatively little is yet known is the so-called *intrinsic factor* required for the absorption of vitamin B_{12} (p. 394).

TABLE 16, 9

Composition of Normal Fasting Gastric Juice (pH 1·0 to 1·5)	
Organic Constituents (0·4 per cent.)	Pepsin Rennin (?) Gastric Lipase Mucin Intrinsic Factor
Inorganic Constituents (0·65 to 0·85 per cent.)	Na^+ K^+ Ca^{++} Mg^{++} $HPO_4^=$ $SO_4^=$ Total chlorides approx. 0·175 N (175 m-equiv./l.) HCl approx. 0·16 N (160 m-equiv./l. or 0·6 per cent. w/v)

When it is secreting vigorously the mucosa becomes red and swollen and its surface covered with profuse wet mucus which is difficult to wash away. Mucin, a glycoprotein (p. 18), is the chief constituent of mucus which, especially if it is present in large amounts, makes the gastric juice thick, sticky and opalescent. Mucus can be secreted independently of hydrochloric acid and pepsin. Mechanical irritation increases the output of mucus which tends to be more viscid and to become cloudy if the irritation is severe. It is able to neutralize considerable amounts of hydrochloric acid and, since it is being formed continuously and remains closely applied to the mucous membrane, it exercises a protective action (see p. 241).

Secretion of hydrochloric acid and pepsin is confined to the *gastric glands* which are found scattered over the cardia and body of the stomach but are absent from the pyloric antrum. They thus occupy some two-thirds of the area of the mucosa (Fig. 16, 4). In histological preparations they appear as parallel tubes often branching near the muscularis mucosæ; on the average four glands open into each pit. The glands possess two different kinds of secretory cells, namely, the *parietal* or *oxyntic* cells concerned in the production of hydrochloric acid and the smaller *peptic cells* or *chief cells* that elaborate pepsinogen, the precursor of pepsin. The former are large wedge-shaped cells staining well with eosin (Fig. 16, 8).


VAGUS N.
LINE OF MUCOSAL DISSECTION INSIDE STOMACH
PYLORIC SPHINCTER
(a)
PYLORIC REGION
INCISION IN GREATER CURVATURE THROUGH ALL LAYERS OF WALL
A
MUCOSA
MUCOSAL PARTITION INSIDE STOMACH
SKIN
B
(b)
VAGUS NERVE
MAIN STOMACH
PARTITION BETWEEN MAIN STOMACH AND POUCH
POUCH
SKIN
SKIN
(c)

SECTION ALONG LINE A– –B IN (b)

FIG. 16, 10

Pavlov's operation for making an isolated pouch of the stomach. Note that the pouch has the same nerve and blood supply as the main part of the stomach.

(*By courtesy of R. A. Gregory.*)

The evidence for the functions of the peptic and oxyntic cells is somewhat indirect. Microchemical analysis of tiny blocks of gastric mucosa shows that acid content and oxyntic cell counts correspond, that is, there are more of both near the neck of the gland (see Fig. 16, 8); pepsinogen is more abundant in the deeper parts. Special stains show that the granules of the peptic cells disappear after vagal stimulation which produces a juice rich in pepsin; histamine which produces an acid secretion poor in enzymes does not affect these granules.

In the gastric mucosa around the entrance of the œsophagus are the *cardiac glands* which secrete mucus only. In the pyloric area of the mucosa the pits are longer and the glands shorter than elsewhere; these *pyloric glands* produce mucus but have no parietal or peptic cells.

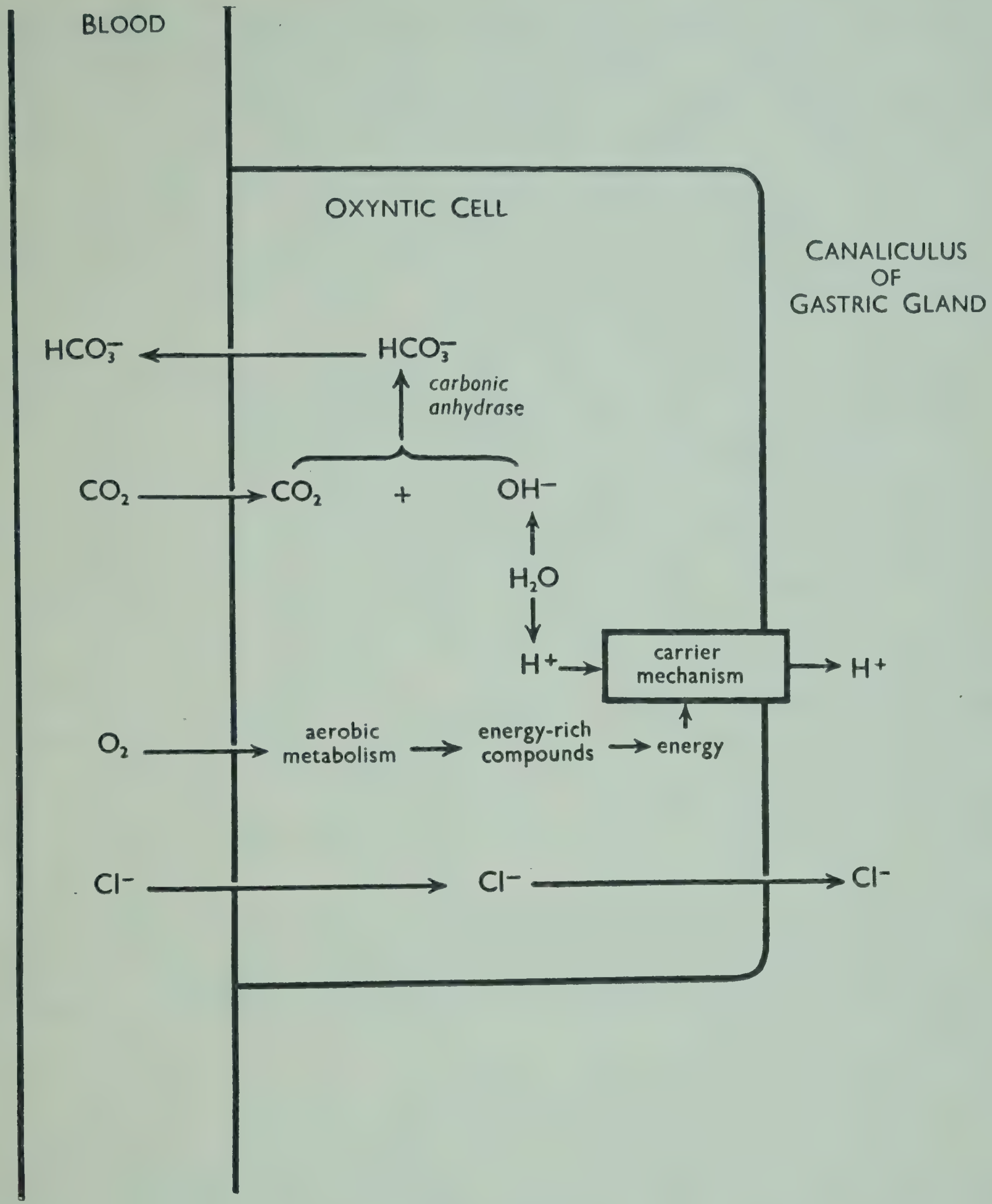

BLOOD
OXYNTIC CELL
CANALICULUS OF GASTRIC GLAND
HCO₃⁻
HCO₃⁻
carbonic anhydrase
CO₂
CO₂
+
OH⁻
H₂O
H⁺
carrier mechanism
H⁺
O₂
aerobic metabolism
energy-rich compounds
energy
Cl⁻
Cl⁻
Cl⁻

FIG. 16, 11

Simplified scheme to illustrate the production of hydrochloric acid in the stomach. The second mechanism (described in the text), involving the production of hydrogen ions from glucose, is not illustrated.

The Hydrochloric Acid of the Gastric Juice.—The oxyntic cells probably produce hydrochloric acid at a concentration of 0·17 N (0·64 per cent.) which has a slightly higher osmotic pressure than the blood but probably the same as the parietal cell cytoplasm. Even in the Pavlov pouch such a high concentration cannot usually be obtained; if, however, the pouch is washed out to remove all mucus and exudate the concentration of hydrochloric acid in the juices approaches 0·16 N. In the intact stomach the acidity is reduced by mucus and other alkaline materials so that the juice collected by a tube contains 0·4 to 0·5 per cent. acid; such a solution is approximately decinormal and has a pH of about 1. This remarkably low pH is, however, not maintained since the acid is constantly being neutralized not only by the food and by the mucin of the saliva and the gastric juice but also from time to time by regurgitation of duodenal contents.

The acidity of the gastric hydrochloric acid represents a concentration of hydrogen ions about a million times greater than is found in plasma. The energy required for this secretion of hydrogen ions against a concentration gradient is derived ultimately from the ærobic oxidation of substrates like glucose, but more immediately from energy-rich phosphate compounds produced by such oxidation (p. 318). Of the many theories which have been proposed to explain the formation of gastric hydrochloric acid those of Davenport and of Davies and Ogston are best known and the following is a greatly simplified account.

The ultimate source of the hydrogen ions of the gastric hydrochloric acid is largely water and to a lesser extent the hydrogen of substrates such as glucose. According to one mechanism postulated by Davies and Ogston hydrogen ions formed by ionization from water are concentrated in an electron-cycle mechanism involving the cytochrome system (p. 83) which transports them across the pericanalicular zone of the oxyntic cells and so into the canaliculus. The energy required for this transport is derived from energy-rich compounds such as creatine phosphate (Ch. 42). At the same time a chloride ion passes from the oxyntic cell into the canaliculus to preserve electrical neutrality (Fig. 16, 11). The hydroxyl ion of water reacts with carbon dioxide to form bicarbonate under the influence of the enzyme carbonic anhydrase which is abundant in the gastric mucosa. These bicarbonate ions then pass into the blood stream (while chloride ions pass in the reverse direction) so that for the secretion of each hydrogen ion into the canaliculus one bicarbonate ion is secreted in the opposite direction into the blood and presumably causes what is usually termed the 'alkaline tide', that is, the slight increase in plasma and urinary pH after a meal. The overall reaction may be represented by the equation

$$2H_2O + 2Cl^- + 2CO_2 + \text{creatine phosphate} \longrightarrow 2H^+ + 2Cl^- + 2HCO_3^- + \text{creatine} + \text{phosphate}.$$

Simultaneously with this process, a second mechanism is thought to operate in which substrates such as glucose are oxidized by removal of hydrogen which is transported by the usual carrier systems (Ch. 8) with the production of reduced flavoproteins. These in turn release hydrogen atoms, the electrons from which are picked up by oxidized cytochrome with the liberation of hydrogen ions.

$$2FPH_2 + 4\ \text{Cyt. Fe}^{+++} \longrightarrow 2FP + 4H^+ + 4\ \text{Cyt. Fe}^{++}.$$

At the same time hydroxyl ions are formed by the reduction of oxygen and react with carbon dioxide to form bicarbonate as explained above.

$$4\ \text{Cyt. Fe}^{++} + O_2 + 2H_2O \longrightarrow 4\ \text{Cyt. Fe}^{+++} + 4OH^-$$

The overall reaction may be illustrated thus:

$$2FPH_2 + O_2 + 2H_2O + 4Cl^- + 4CO_2 \longrightarrow 4H^+ + 4Cl^- + 2FP + 4HCO_3^-.$$

ENZYMES OF THE GASTRIC JUICE

The proteolytic enzyme of the stomach, *pepsin*, is secreted by the peptic cells of the fundus in the form of an inactive precursor or zymogen called *pepsinogen* which, on coming into contact with hydrochloric acid, is converted into pepsin. Pepsin is an unusual enzyme in that its optimum pH value is in the neighbourhood of 1·2, and it therefore meets conditions most suited to its action in the acid contents of the stomach. Pepsin breaks down protein to the stage of proteoses or peptones, that is to say, it breaks the protein chain into shorter fragments, but it does not liberate free amino-acids. The products of peptic digestion are later broken down to amino-acids by the proteolytic enzymes of the pancreas and intestinal juice.

The urine in man contains uropepsinogen, apparently identical with gastric pepsinogen. The enzyme must be derived from the gastric glands since it disappears after gastrectomy. Many patients with duodenal ulcer eliminate in the urine more than normal amounts of uropepsinogen.

The stomach of the calf, and of other young animals, contains an enzyme named *rennin* which brings about the clotting of milk. This enzyme is present in commercial rennet used for making junket (curds). Purified rennin possesses only slight proteolytic activity, whereas pepsin itself, even in the crystalline form, is able to bring about the clotting of milk. The action of rennin on caseinogen has been discussed on p. 173. The optimum pH for rennin is 5 or 6, a value which corresponds to the reaction normally said to be found in the stomach of a healthy infant, although recent evidence indicates that the acidity may be considerably greater than this. It is not certain that rennin is present in the gastric juice in the adult but, if it is, it must be almost inactive at the low pH prevailing there.

Gastric *lipase* is a fat-splitting enzyme of little importance.

In spite of the presence of high concentrations of hydrochloric acid as well as powerful protein-splitting enzymes the surface epithelium of the stomach and intestine retains its integrity during life although it is rapidly digested after death. Not all areas of the alimentary tract are tolerant of these powerful agents and it appears that parts like the stomach, duodenum and transverse colon, which can secrete large amounts of mucus, are more resistant than others like the jejunum. The protective action of mucus cannot be due simply to the buffering action of the protein mucin but the recent finding that mucin can withstand the action of pepsin and inhibit its activity may provide a clue to the mechanism of the process. Anti-enzymes (p. 75) may also contribute to the resistance of the epithelium.

CONTROL OF GASTRIC SECRETION

Nervous Mechanism.—The secretion of gastric juice is partly under the control of the nervous system by way of the secretory fibres in the vagus nerves. Stimulation of the peripheral end of a cut vagus nerve is followed after a latent period of some minutes by the secretion of a highly acid juice containing much pepsin; after prolonged vagal stimulation nearly all the pepsinogen granules disappear from the peptic cells.

Acetylcholine is liberated at the vagal nerve endings and stimulates all the secretory cells directly except the oxyntic cells. For the latter there is a double humoral mechanism by which acetylcholine liberates histamine which then stimulates the oxyntic cells to produce hydrochloric acid; some of this histamine escapes into the gastric juice. The importance of the vagus nerves for gastric secretion was demonstrated by Pavlov by experiments on dogs in which the œsophagus had been cut in the neck and the two ends brought to the surface of the skin as separate openings. Food which the animal chewed and swallowed was discharged to the exterior instead of entering the stomach. Under such conditions of 'sham feeding' a copious secretion of gastric juice was produced when the vagi were intact, but when the nerves were cut the secretion was absent.

In animals stimulation of the sympathetic supply to the stomach causes vasoconstriction but the effect on secretion is variable. Intravenous infusion of adrenaline or noradrenaline inhibits histamine-induced secretion; this is presumably not a direct effect but simply the result of the reduction in blood supply caused by vasoconstriction. Coeliac ganglionectomy has no effect on the secretion of the dog's stomach. There is no morphological evidence of a sympathetic supply to the gastric mucosa and it seems likely, as Pavlov asserted sixty years ago, that the vagus is the only nerve concerned in gastric secretion.

Pavlov showed that it is possible in dogs to build up conditioned reflexes (Chapter 50) as a result of which gastric secretion is produced by a great variety of stimuli. Thus the sight or smell of food can initiate a flow of gastric juice, especially in hungry animals. Such secretion has been called *psychic* or *appetite juice*. The secretion of gastric juice at the beginning of a meal, of considerable importance in the digestive process in man as well as in the dog, is sometimes referred to as the *reflex* or *cephalic phase* of gastric secretion (Fig. 16, 12); it is readily inhibited by emotions such as fear and disgust and probably by other psychic influences.

Chemical Mechanisms.—The presence of food in the stomach is itself capable of provoking the continued secretion of gastric juice; this is the so-called *gastric phase* of gastric secretion. It is probable that the further digestion of the meal is maintained by this mechanism which is independent of the nervous system. Certain articles of food, especially meat, and soups made from meat, are particularly active as secretagogues, as are also some of the products of protein digestion. It is now believed that these materials exert their action indirectly by causing the liberation of *gastrin* from the mucosa of the pyloric antrum. This hormone is carried in the blood to the gastric glands and maintains the secretion of gastric juice initiated by the nervous mechanism. Gastrin was for a time thought to be identical with histamine, itself a powerful stimulator of gastric secretion, but potent extracts, prepared from the mucous membrane of the stomach or small intestine, containing protein but free from histamine, have now been obtained. Such materials may act by liberating histamine in the gastric mucosa. The juice secreted in response to the hormone is greatly inferior in peptic activity to that produced by stimulation of the vagus.

The gastric glands are also stimulated to secrete when a number of substances, such as the breakdown products of the gastric digestion of proteins, leave the stomach and come into contact with the duodenal mucosa. This stimulatory effect is probably brought about by the liberation of gastrin which is carried from the duodenum to the mucosa in the blood (*intestinal phase* of gastric secretion). The effect is, however, much less marked than it is in the pyloric part of the stomach.

These conclusions about the hormonal control of gastric juice are founded on experiments such as those illustrated in Figs. 16, 13 and 16, 14, in which pouches of the whole stomach or of part of the stomach of dogs were made. Fig. 16, 13 A shows the daily output of hydrochloric acid from a Pavlov pouch. If two-thirds of the antrum were resected the acid output was a little less (B) but complete removal of the antrum reduced the acid output to one-sixth (C). Since the nervous connexions to the pouch were meagre the gastric phase of secretion seems to be about six times more important than the intestinal. If a pouch of the whole stomach with the vagus nerves cut is made as in Fig. 16, 14 B the food cannot come into contact with the antrum and the acid output is quite small. Reimplantation of the antrum, but not of any other part of the stomach, into the duodenum (Fig. 16, 14 C), or even the colon, restores the secretion of hydrochloric acid to its former level. On the basis of such experiments Dragstedt believes that in the dog the cephalic phase and the gastric phase each account for about 45 per cent. of the gastric juice and the intestinal phase for 10 per cent. or less.

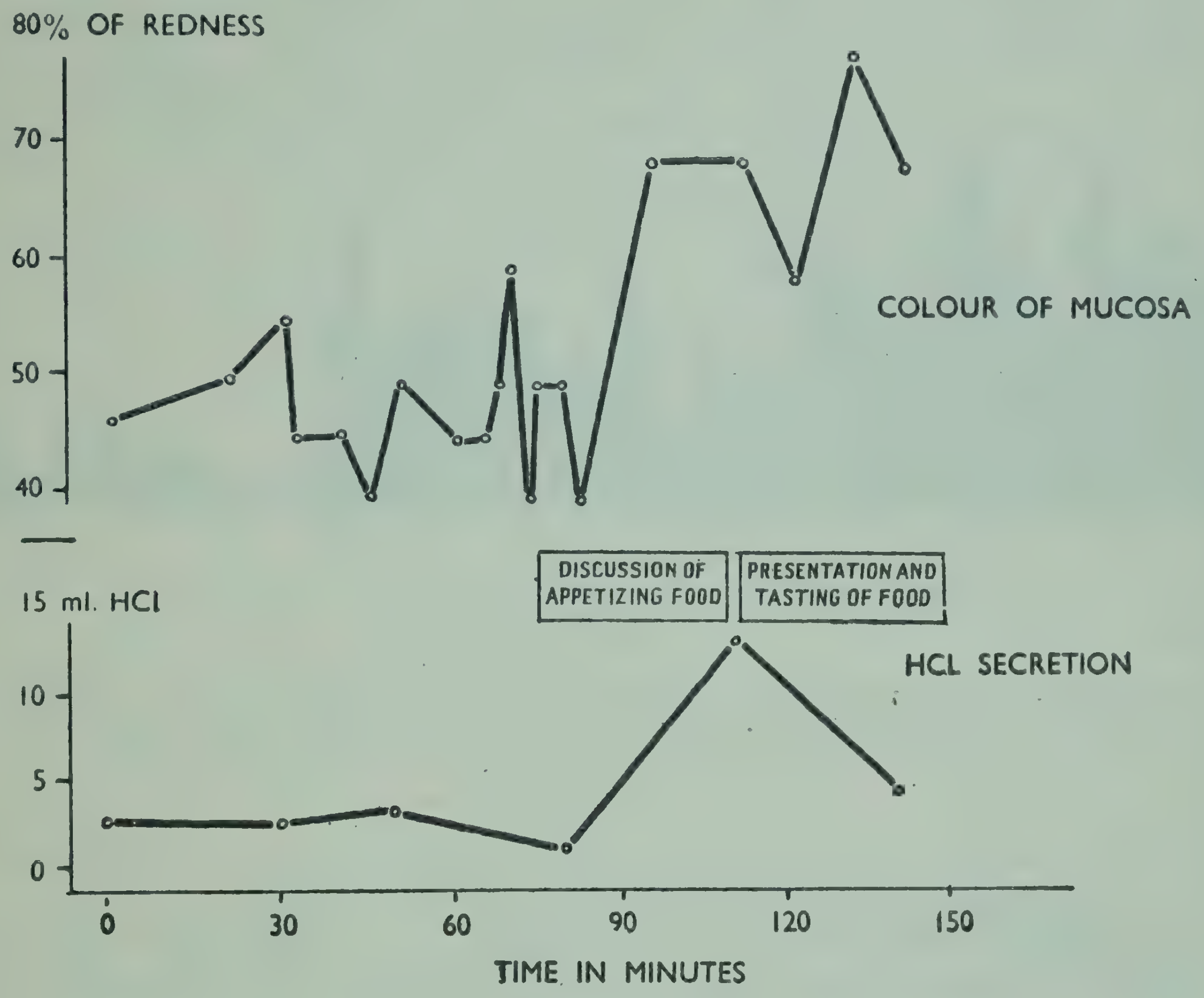


Fig. 16, 12

Response of the human stomach (Tom) in terms of vascularity and acid secretion to stimulation of appetite. Acid secretion is expressed as ml. of 0·1 N HCl per 15 min. Mucosal colour: 90 = red, 50 = pale. (S. Wolf & H. G. Wolff (1947), *Human Gastric Function*. Oxford University Press.)

In an experiment in which completely denervated pouches of the body of the stomach and pylorus had been made, distension of the pyloric pouch caused secretion in the other pouch. The response in the body pouch could not be due to absorption of secretagogues and must have been brought

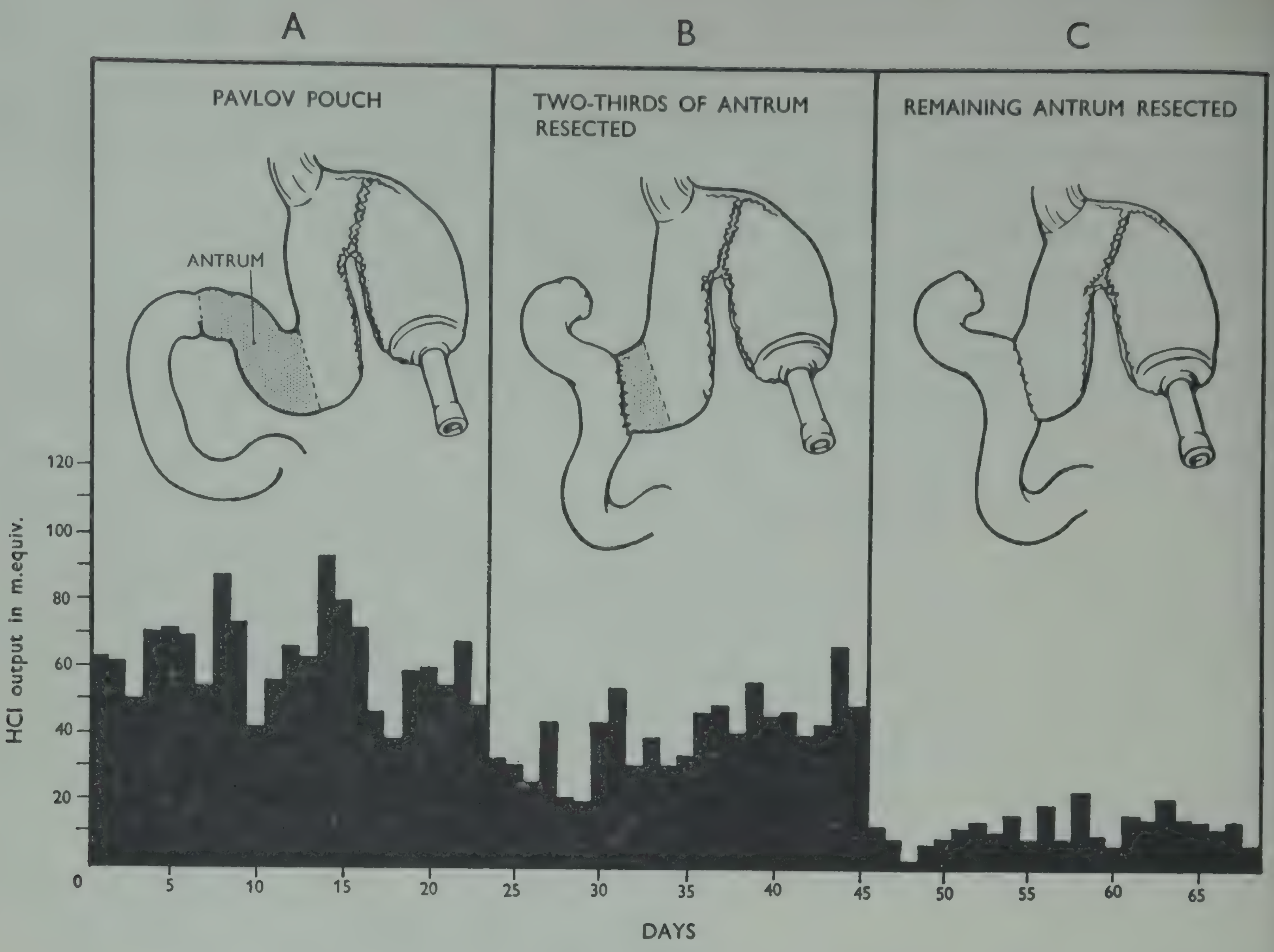


FIG. 16, 13

Effect of partial removal of antrum on gastric secretion in Pavlov pouch. L. R. Dragstedt, E. R. Woodward, E. H. Storer, H. A. Oberhelman & C. A. Smith (1950), *Ann. Surg.* **132,** 629.

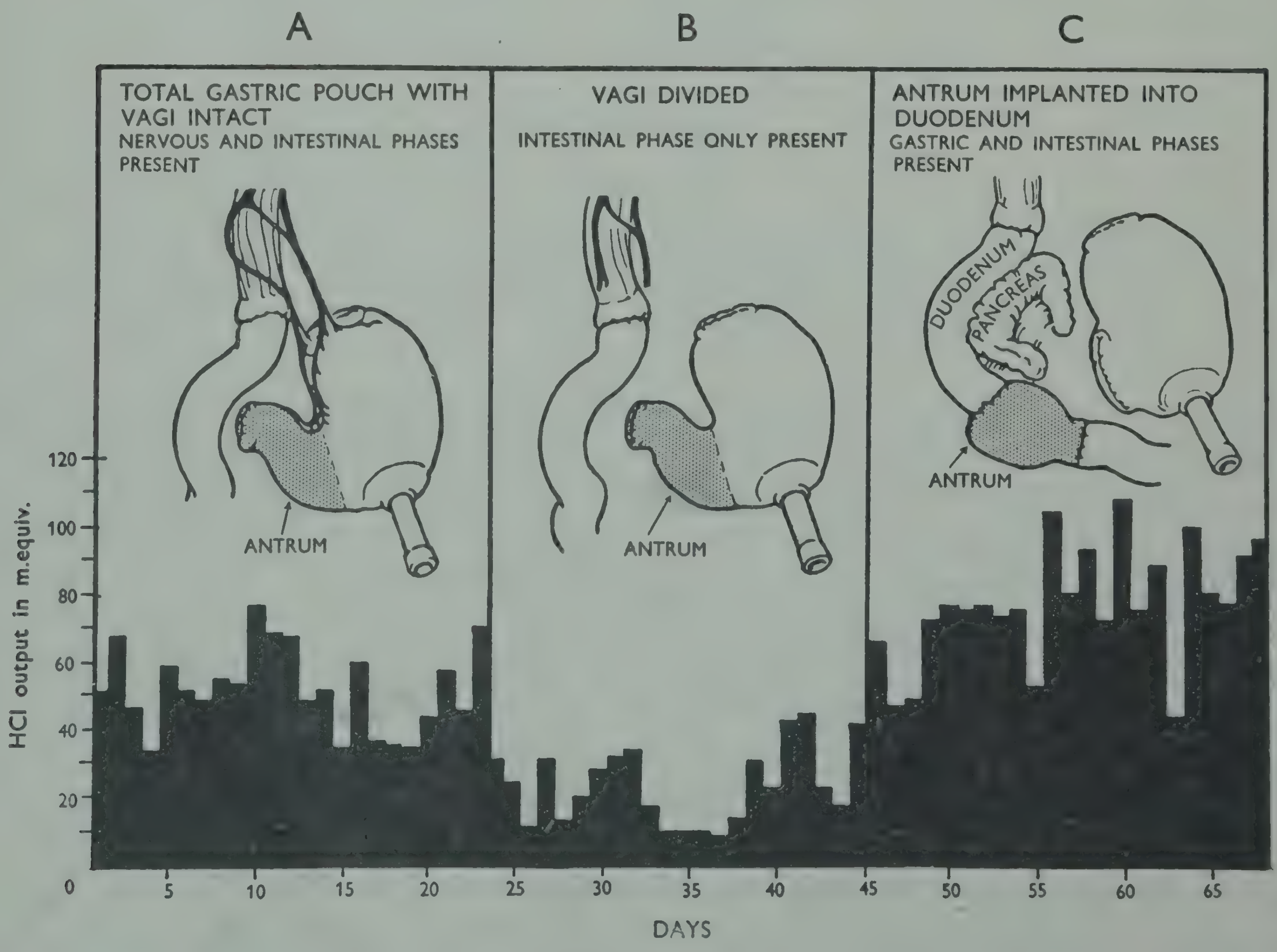


FIG. 16, 14

Relative importance of the three phases of gastric secretion. L. R. Dragstedt, E. R. Woodward, E. H. Storer, H. A. Oberhelman & C. A. Smith (1950), *Ann. Surg.* **132,** 635.

about by a hormonal mechanism. Acetylcholine placed in the pyloric mucosa caused secretory hormones to be released into the circulation and this effect like that of distension was prevented by atropine. It is thus probable that gastrin is released by a local cholinergic mechanism. In man the amount of secretion is affected by the volume of the meal; increase in distension up to a certain point increases the secretion but greater distension reduces it.

Similar work has shown that the character of the juice produced by this hormonal action is dependent on the kind of food in the stomach. Meat produces the greatest volume of juice, while the greatest peptic power is found when bread is given.

Gastric Inhibitors.—Gastric digestion may be inhibited to some extent by fat. This action is apparently not due to the direct action of fat on the gastric mucosa but to the liberation of a hormone which acts as a gastric inhibitor, since, when a fat meal is given the motility and acid secretion of a denervated pouch is inhibited as the fat meal passes into the duodenum. Presumably this is the basis for the popular opinion that fats are 'indigestible'. Olive oil is often given to patients with gastric or with duodenal ulcer in order to inhibit the secretion of acid. Powerful inhibition of gastric secretion, especially of hydrochloric acid, may be produced by the substance enterogastrone, which Ivy and his colleagues have prepared in purified form from extracts of upper intestinal mucosa. Its physiological significance is not yet clear.

Contrary to the usual belief, there is no reduction in acid secretion, either spontaneous or histamine-induced, in pregnancy. Administration of progesterone can reduce secretion slightly.

To summarize: Apart from the very important psychic factors which often initiate secretion a meal provokes gastric secretion in two main phases: (1) cephalic, which begins within a few minutes of taking the food and which depends on the integrity of the nervous pathways; (2) a humoral or gastric phase, which begins much later and keeps up the secretion already begun in the first phase. This is due to the indirect action of the products of digestion in the stomach. A third or intestinal phase may be added, due to the action of the products of digestion in the duodenum. Distension also plays a small part in eliciting secretion. The first two phases follow one another without any break, the pepsin content being under mainly nervous control whereas the acid and water content are mainly under hormonal control.

DIGESTION IN THE STOMACH

The stomach is not an essential organ and animals from which it has been removed or human subjects from whom it has been excised because of disease continue to live without much discomfort, although frequent small meals are essential at first. It may be necessary to make good the loss of the intrinsic factor required for absorption of vitamin B_{12} (p. 394). In the healthy human being the function of the stomach is essentially to reduce the food to a semifluid mass of uniform consistency called *chyme*, which is passed on in controlled quantities to the duodenum where the main digestive processes begin.

While the food remains in the stomach the action of salivary ptyalin continues until it is arrested by the penetration of hydrochloric acid throughout the food mass. Proteins are broken down to proteoses and peptones by the action of pepsin. The concentration of hydrochloric acid in the stomach is great enough to bring about some hydrolysis of sugars containing fructofuranose units, for example sucrose. Hydrochloric acid also limits the growth of micro-organisms entering the stomach. Fats are not hydrolysed to any appreciable extent in the stomach but they are intimately mixed with the food mass to form a coarse emulsion.

Gastric Analysis

The gastric mucosa secretes throughout the 24 hours and produces in that time some 2 to 3 litres of juice but the rate of secretion is, as we have seen, not constant. The fasting stomach contains resting juice which does not exceed 50 to 100 ml. The effect of food is to neutralize such acid as is present in the resting juice (Fig. 16, 15) and to increase the secretion of acid.

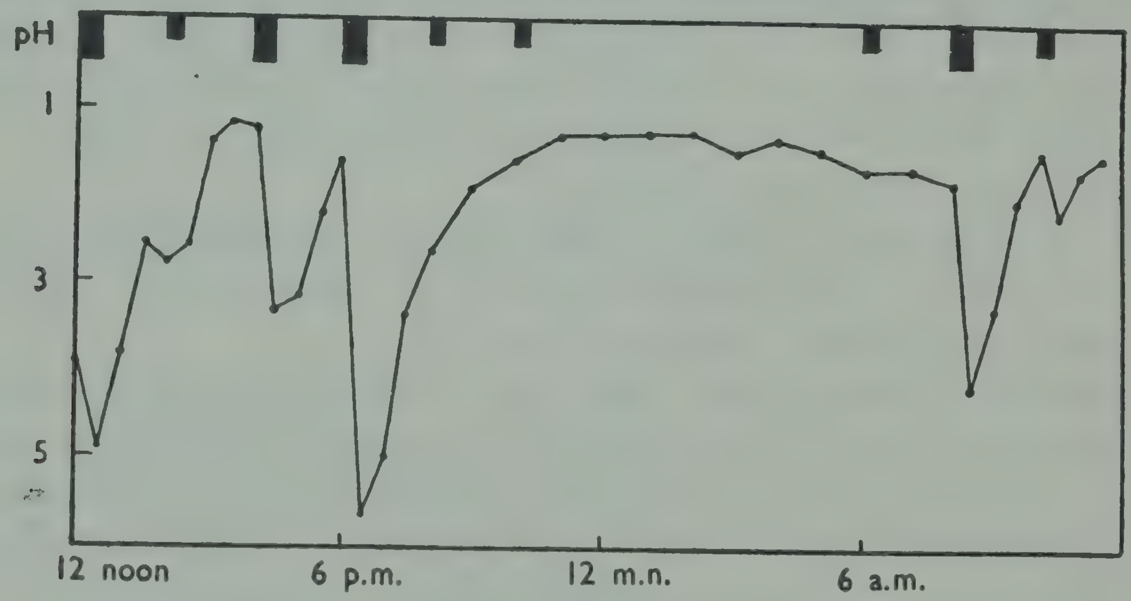


Fig. 16, 15

The pattern of gastric acid as shown by a healthy male subject, aged 36. The pH of samples of gastric juice obtained by a Ryle's tube lying in the stomach was measured by a glass electrode. The large black rectangles indicate meals, the small ones drinks of milk (A. H. James & G. W. Pickering (1949). *Clin. Sci.* **8**, 181).

In the study of gastro-intestinal disorders it has in the past been customary to study the pattern of gastric acidity following the taking of a standard meal, for example a pint of thin gruel, by the process of *fractional gastric analysis*. This consists in withdrawing small samples of the contents of the stomach before and at quarter or half-hourly intervals after the meal and determining the concentration of acid in them. However, the figures obtained are affected by so many variables such as the position of the tube, the amount of salivation, regurgitation from the duodenum and the physical and mental state of the subject, that they are virtually impossible of interpretation. To illustrate this it can be stated that a high concentration of hydrochloric acid in the juice does not indicate hypersecretion. As the oxyntic cells secrete acid always at a concentration of 0·17 N the concentration in the juice depends mainly on two variables—the rate of emptying and the rate of secretion. For example if the rate of secretion is high and there is little fluid in the stomach to dilute it, because of rapid emptying, the level of acid in the juice is high. Furthermore, variations in the acidity of gastric juice obtained from healthy persons are considerable and it is probable that differences in secretory activity occur from time to time even in the same individual.

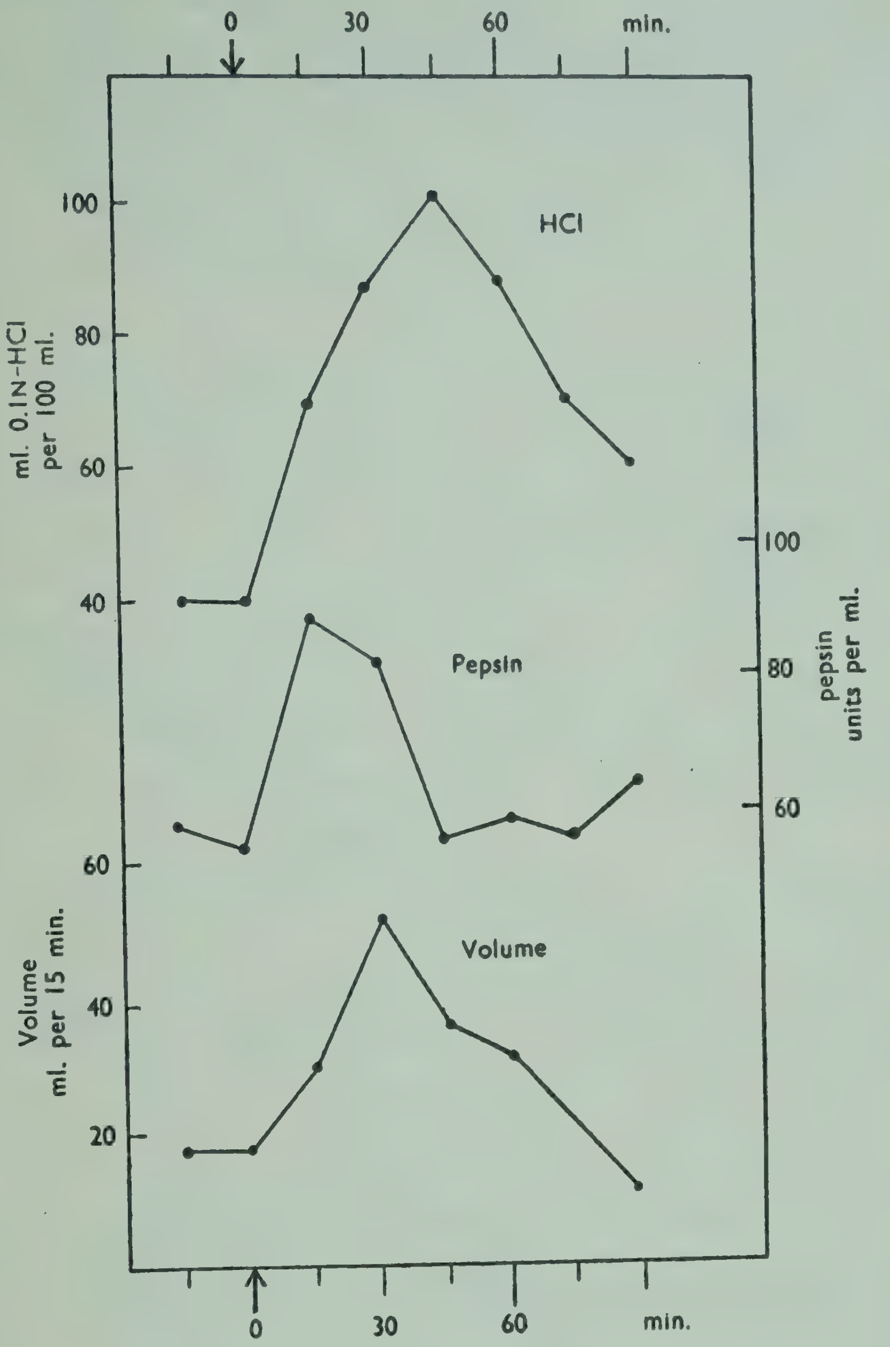


FIG. 16, 16

The effect of injection of histamine on gastric secretion in man. Mean results in 14 subjects. 0·54 mg. of histamine was injected at the arrow marked 0. A fasting sample was taken 15 min before. The gastric contents were aspirated continuously and the volume at the end of each 15 min period was recorded. Although there was only a moderate rise in pepsin concentration the concurrent rise in volume of juice shows that the pepsin output was considerably increased. In fact, at ½ hr after injection of histamine the output rate was four times that in the fasting condition. (C. A. Ashford, H. Heller & G. A. Smart (1949), *Brit. J. Pharmacol.* **4**, 153.)

More useful information is obtained by determining the response of the gastric mucosa to histamine (Fig. 16, 16). It is usually sufficient to aspirate a sample of resting juice and further samples 30 minutes after each of three successive doses of 0·5 mg. of histamine phosphate, given by subcutaneous injection. The pH of each specimen is then assessed with Töpfer's reagent which changes colour at pH 3·5; should the pH of any specimen be below 3·5 it is taken that the gastric mucosa is capable of secreting acid in response to stimulation. Much larger doses of histamine

(0·04 mg./kg. body weight) in conjunction with an antihistamine drug are, however, necessary to obtain maximum and consistent output of hydrochloric acid. Since in many cases it is sufficient to know simply whether the stomach secretes hydrochloric acid or not, the ingenious procedure devised by Segal is useful, especially as it eliminates the discomfort of swallowing a stomach tube. The patient takes a quantity of quininium resin and is given histamine subsutaneously. At pH 4 or below hydrogen ions displace the quininium cation and the quinine formed passes from the stomach into the small intestine where it is absorbed and excreted by the kidneys. The quinine can be estimated in the urine by its fluorescence in ultraviolet light. If certain ions such as aluminium and iron are excluded form the diet for several days beforehand, the excretion of 30 μg. or over of quinine in two hours indicates the presence of free acid in the stomach.

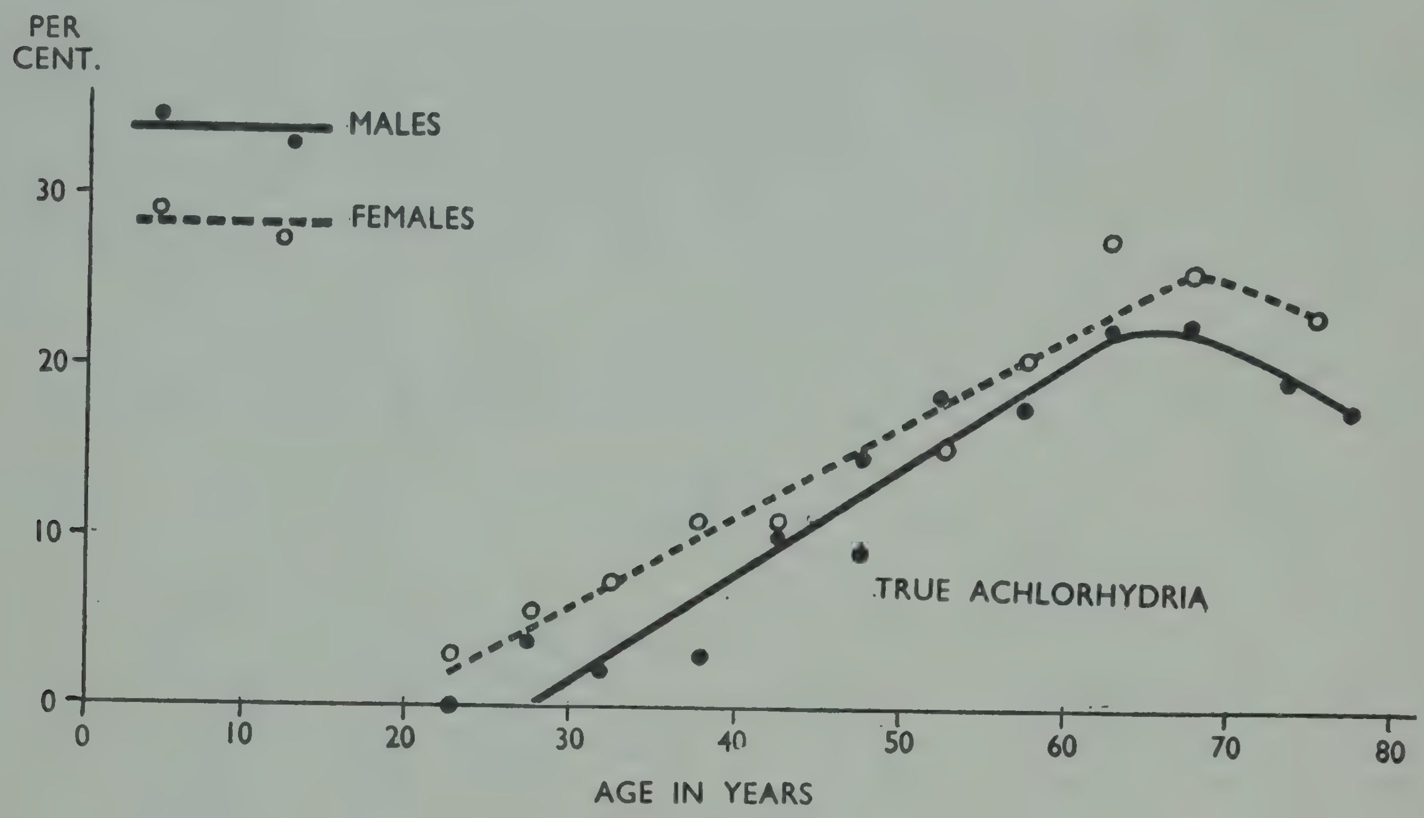


Fig. 16, 17

Incidence of achlorhydria in 3746 persons.
(*After* F. R. Vanzant, W. C. Alvarez, G. B. Eusterman, H. L. Dunn & J. Berkson (1932), *Arch. intern. Med.* **49**, 349.)

Quite a number of people show no free hydrochloric acid in the juice after a gruel meal and a small percentage (Fig. 16, 17) of apparently healthy adults may have 'true achlorhydria,' that is the inability to secrete acid in response to an injection of histamine, but the incidence of achlorhydria in the general (healthy) population has probably been exaggerated owing to failure to verify radiologically that the tube is in fact in the stomach. If large enough doses of histamine, covered with antihistamine drugs, are given the number of persons in whom the pH would not fall to 3·5 or less is probably very small.

When a small dose of soluble insulin (for example, 7 units) is given subcutaneously to a man there is, after a latent period of about 40 minutes during which blood sugar falls to about 70 mg. per 100 ml., a copious secretion of gastric juice. This is not a direct effect of insulin but is dependent on the hypoglycæmia, because there is no secretion of juice when glucose is given with the insulin to maintain the blood-sugar level. Since this effect of insulin on gastric secretion is abolished by double vagotomy, or by large doses of atropine, the hypoglycæmia must act centrally, probably on the hypothalamus, and the administration of insulin can be regarded as a method of stimulating the mucosa via the vagus nerves. The secretion produced contains both hydrochloric acid and pepsin just as in animal experiments in which the vagus is stimulated directly.

The administration of insulin followed by histamine produces a very marked flow of gastric juice and presumably this might be used as a test for estimating the maximum secretory powers of the stomach.

VOMITING

Vomiting is usually accompanied by nausea and preceded by retching, in which the stomach is periodically compressed by a series of more or less violent contractions of the diaphragm and the abdominal muscles. Along with nausea and vomiting there are autonomic disturbances such as pallor, slowing of the heart, sweating and salivation. Vomiting may arise from a great variety of causes, including such diverse and unrelated phenomena as feelings of disgust, severe pain and increased intracranial pressure. Some individuals vomit much more readily than others and one person vomiting as a result of sea-sickness or whooping-cough may influence observers to follow his example.

If vomiting has to be initiated as a first-aid measure to empty the stomach, irritating the back of the throat, or the drinking of warm salt-and-water or mustard-and-water may be effective. Such preparations are said to be *emetics.* The threshold dose of the emetic substance copper sulphate given by mouth to cats is increased two-fold by section of the vagus nerves but eight-fold by complete denervation of the stomach. Sympathetic and parasympathetic nerves must therefore both carry afferent impulses from the gastric mucosa irritated by copper sulphate. After complete denervation of the stomach, copper produces vomiting only after a long latent period which is the time taken for the absorption of copper into the blood from the gastrointestinal tract; once in the blood stream it acts on the brain stem.

The threshold dose of the emetic drug apomorphine administered intravenously to cats is unaffected either by section of the vagus nerves or by complete denervation of the stomach. This suggests that it is acting on an *emetic centre* in the brain. In the decerebrate cat Wang and Borison have found that electrical stimulation of a small area on the dorsolateral border of the reticular formation of the medulla, which includes the tractus solitarius and its nucleus, produces vomiting. After destruction of this area on both sides vomiting cannot be induced either by apomorphine given intravenously or copper sulphate given by mouth.

The act of vomiting is brought about by a complex series of movements in which antiperistalsis in the stomach plays little, if any, part. Indeed vomiting can be produced in animals in which the stomach has been replaced by a rubber bag. According to X-ray observations the stomach contracts at the angular notch and the contents of the pyloric antrum are forced into the relaxed body and fundus. A deep inspiration is taken and the glottis closes; the soft palate is raised to prevent vomitus entering the nose. A sharp increase in intra-abdominal pressure, brought about by a violent contraction of the abdominal muscles, forces the gastric contents through the relaxed cardia and along the œsophagus into the mouth. In an unconscious subject the glottis may not close completely and vomited material may be inhaled into the lungs.

REFERENCES

BABKIN, B. P. (1950). *Secretory Mechanism of the Digestive Glands.* New York: Hoeber.

BOCKUS, H. L. (1947). *Gastroenterology,* Vol. I. London: W. B. Saunders Co.

CODE, C. F., HIGHTOWER, N. C. & MORLOCK, C. G. (1952). Motility of the alimentary canal in man. *Amer. J. med.* **13,** 328–351.

CONWAY, E. J. (1953). *The Biochemistry of Gastric Acid Secretion.* Springfield, Illinois: Thomas.

DAVIES, R. E. (1951). The mechanism of hydrochloric acid production by the stomach. *Biol. Rev.* **26,** 87–120.

GROSSMAN, M. I. (1950). Gastrointestinal hormones. *Physiol. Rev.* **30,** 33–90.

HEINZ, E. & ÖBRINK, K. J. (1954). Acid formation and acidity control in the stomach. *Physiol. Rev.* **34,** 643–673.

HUNT, J. N. (1952). Gastric secretion: assessment. In *Modern Trends in Gastroenterology,* pp. 296–308. London: Butterworth.

ILLINGWORTH, C. F. W. (1950). The acid secretion of the stomach. *Edinb. med. J.* **57,** 225–233.

JAMES, A. H. (1956). *Gastric Function.* Monograph No. 4 of the Physiological Society. London: Arnold.

KAREL, L. (1948). Gastric absorption. *Physiol. Rev.* **28,** 433–450.

WANG, S. C. & BORISON, H. L. (1950). The vomiting centre. *Arch. Neurol. Psychiat. Chicago,* **63,** 928–941.

WOLF, S. & WOLFF, H. G. (1943). *Human Gastric Function.* Oxford University Press.

YOUMANS, W. B. (1952). Neural regulation of gastric and intestinal motility. *Amer. J. Med.* **13,** 209–226.

CHAPTER 17

DIGESTION AND ABSORPTION IN THE SMALL INTESTINE

IN this chapter the manifold processes involved in digestion and absorption are discussed. As the pancreas is perhaps the most important source of digestive juices it is desirable to consider it first.

THE PANCREAS

The pancreas is remarkable in producing both an external and an internal secretion. The bulk of the gland has a racemose structure resembling that of the salivary glands (Fig. 17, 3). Two zones can be recognized in the acinar cells, that nearer the lumen containing zymogen granules which are discharged during periods of secretory activity, and a peripheral clearer zone (Fig. 17, 4). The main duct of the gland communicates with the bile duct at the *ampulla of the bile duct* (ampulla of Vater) and discharges the secretion into the duodenum through the

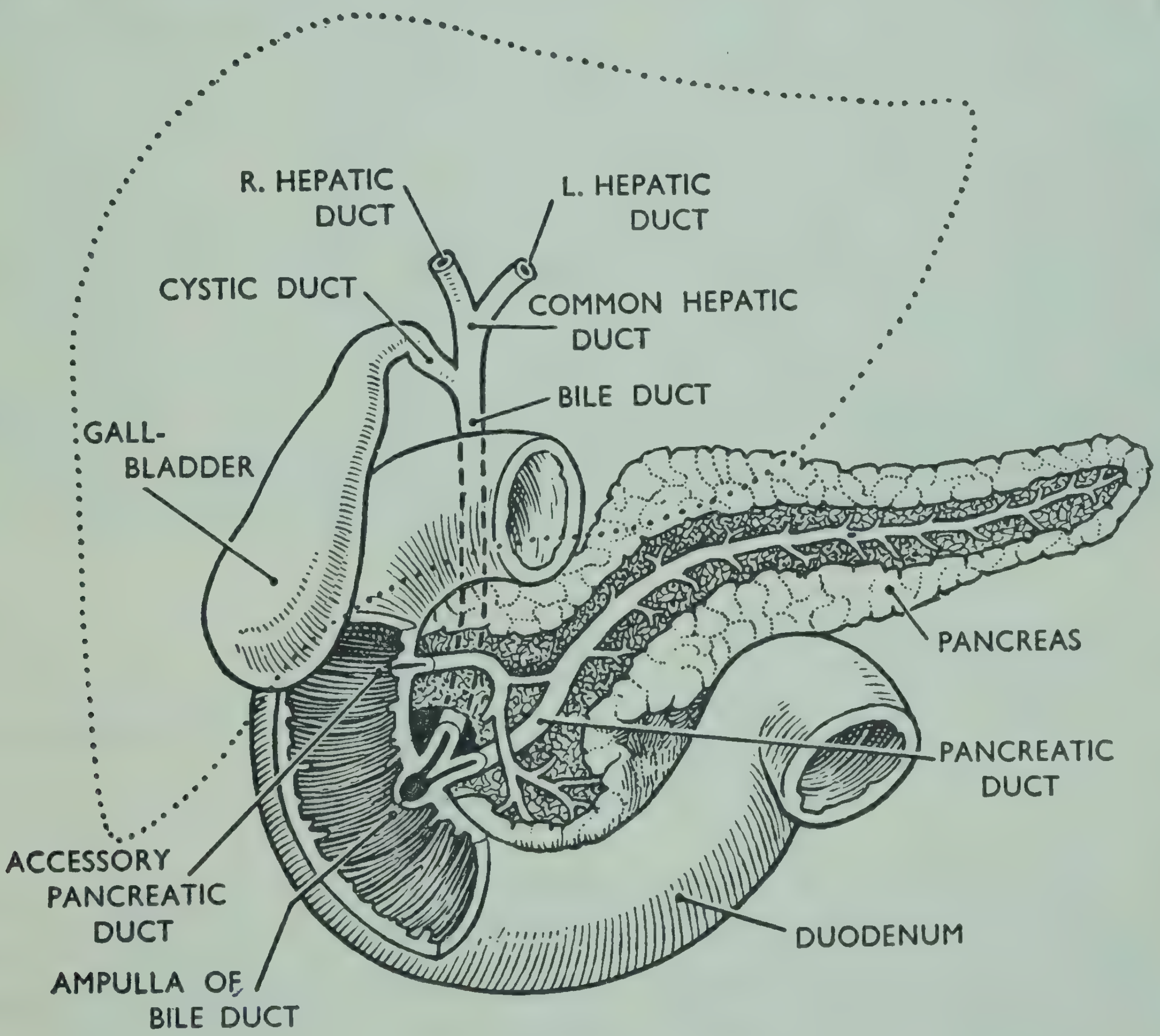

R. HEPATIC DUCT
L. HEPATIC DUCT
CYSTIC DUCT
COMMON HEPATIC DUCT
BILE DUCT
GALL-BLADDER
PANCREAS
PANCREATIC DUCT
ACCESSORY PANCREATIC DUCT
DUODENUM
AMPULLA OF BILE DUCT

FIG. 17, 1

Connexions of the liver, gall-bladder, and pancreas, with the duodenum. The anatomical arrangement of the entry of the pancreatic duct and the bile duct is not always as shown here; they enter the duodenum separately in 29 per cent of persons.

sphincter of Oddi (Fig. 17, 1). Embedded within the pancreatic acinar tissue is a relatively small amount of cellular tissue, quite different in appearance, which occurs in little groups, the *interalveolar cell-islets* or the islands of Langerhans. This tissue produces the internal secretion, *insulin*, which is described in Chapter 51.

The pancreatic secretion may be collected from animals by inserting a cannula into the main duct, or by mobilizing a portion of the duodenum containing the opening of the duct, and stitching it to the surface of the abdominal wall.

In man the amount of juice collected from a fistula per day is 1200 to 1500 ml. It is a clear colourless fluid, with a pH of about 8, a specific gravity of 1·008 to 1·030 and has the same osmotic pressure as plasma. It contains about 1 per cent. of inorganic materials, the most important being sodium bicarbonate, which gives it a degree of alkalinity such that one volume of pancreatic juice is capable of neutralizing an equal volume of acid gastric content obtained simultaneously from the same animal. When the acid gastric contents pass through the pylorus into the duodenum they are neutralized by the bicarbonate of the pancreatic secretion. Pancreatic juice contains 0·1 to 0·3 per cent. of organic material, mainly enzymes, the chief of which is *trypsin*. This enzyme is secreted in the form of an inactive precursor, *trypsinogen*, which is converted into active trypsin by *enterokinase* secreted by the duodenal mucosa. In the presence of trypsin, trypsinogen may also be activated autocatalytically. Trypsin which has

TABLE 17, 2

Composition of Pancreatic Juice (pH 7·1 to 8·2)	
Organic constituents (0·1 to 0·3 per cent.)	Trypsinogen Chymotrypsinogen Carboxypeptidase Aminopeptidase Lipase Amylase Maltase Sucrase Lactase Nucleases Albumin and globulin
Inorganic constituents (1·0 per cent.)	Na^+ K^+ Ca^{++} Mg^{++} Cl^- $SO_4^=$ $HPO_4^=$ HCO_3^- (0·5 to 0·7 per cent.)

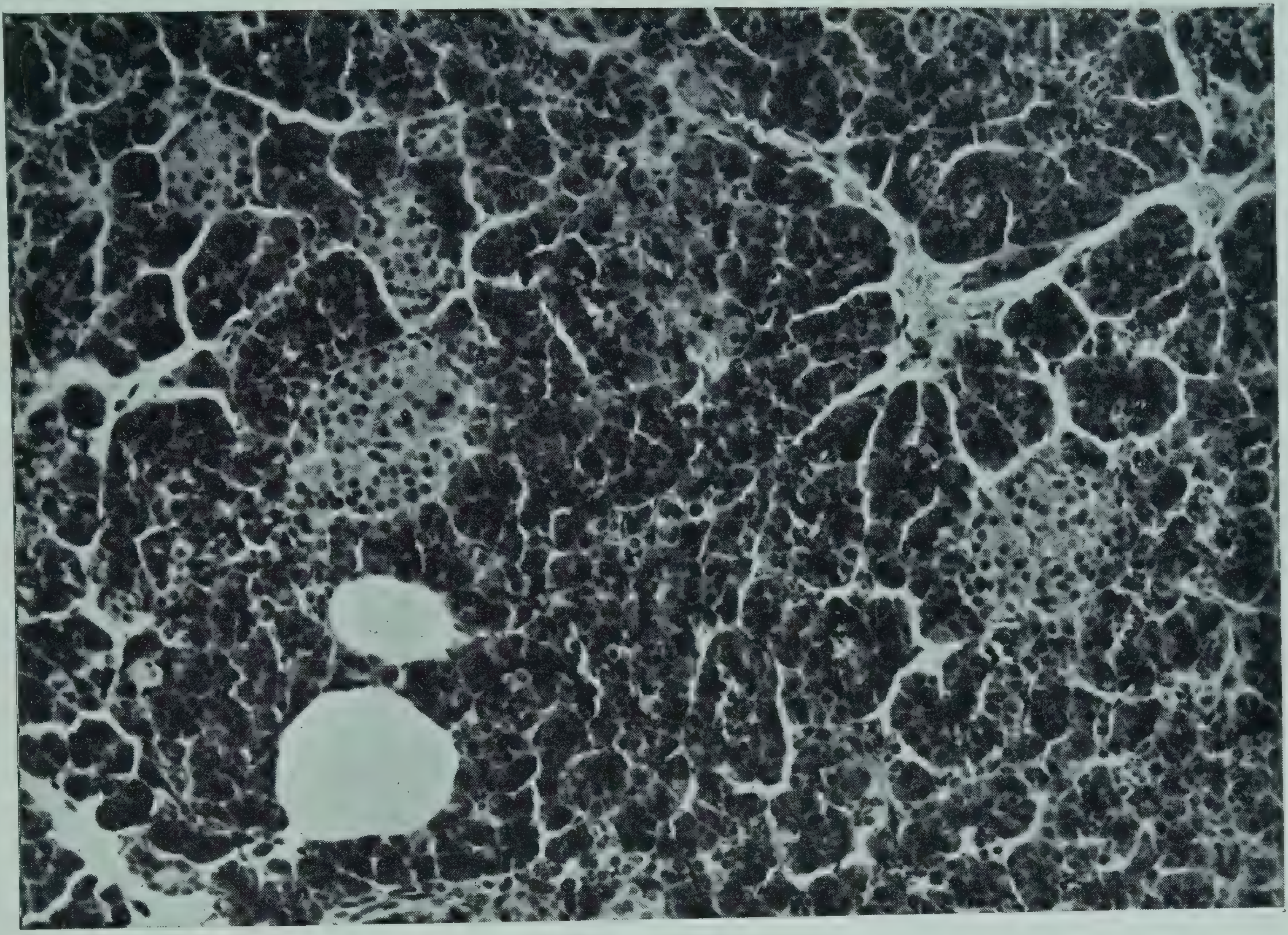


FIG. 17, 3

A section through a human pancreas. The acinar tissue is deeply stained, whereas the islet tissue is pale with dark nuclei. The section shows two circular spaces originally containing fat and also a ramifying connective tissue framework. Magnification ×120. Stain: hæmalum, phloxin and tartrazine.

(*By courtesy of A. C. Lendrum.*)

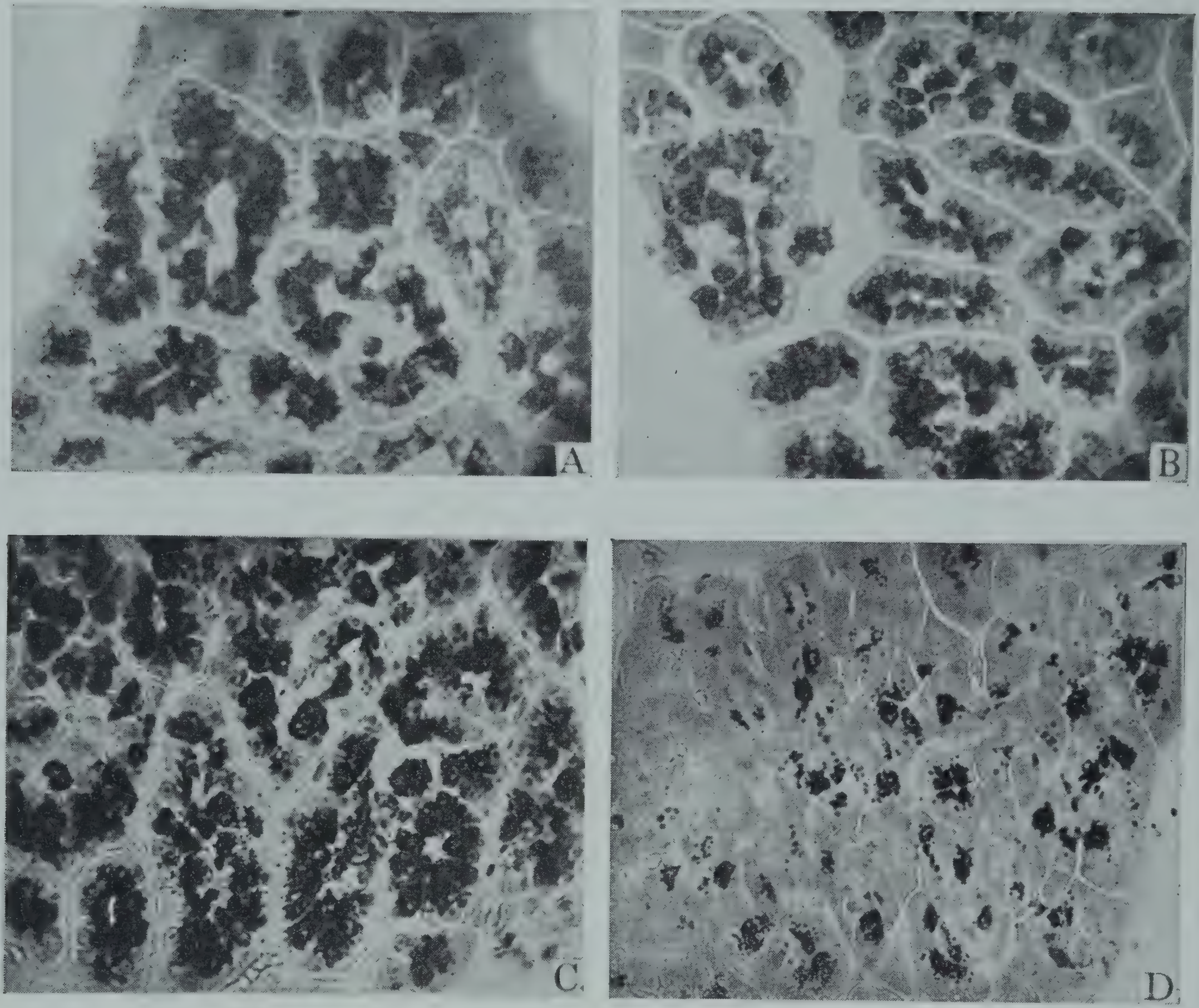


FIG. 17, 4

A. Section of pancreas stained to show granules. This is the resting appearance before injection of secretin. Control for B.
B. Same pancreas after injection of secretin. No loss of granules. Juice was of low enzyme content.
C. Control for D.
D. Same pancreas after injection of pancreozymin, which produced a juice rich in enzymes. Note loss of granules.

Magnification in all cases × 500.

(A. A. Harper & I. F. S. Mackay (1948), *J. Physiol.* **107**, 96.)

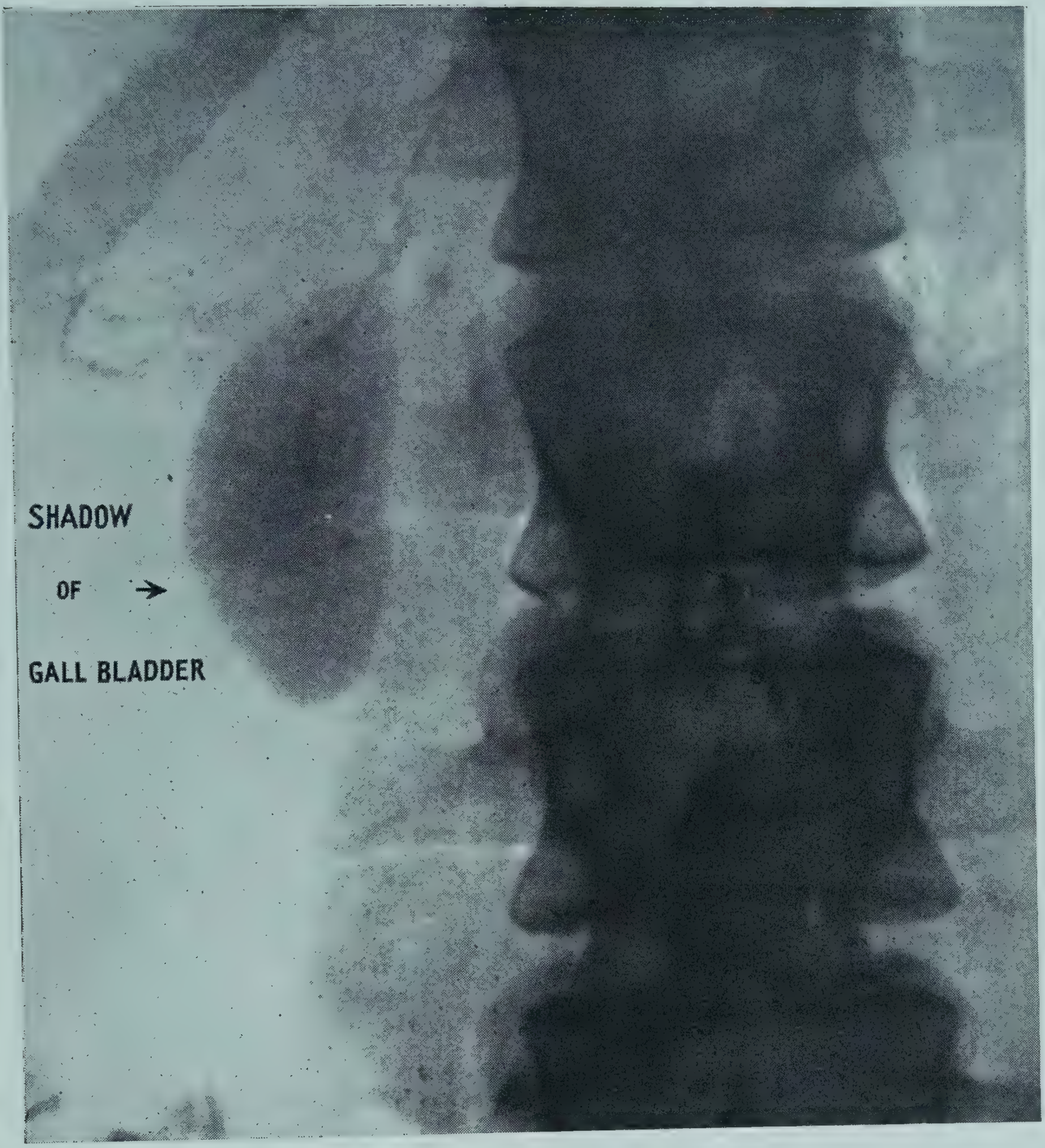


Fig. 17, 5

X-ray photograph of the right subcostal region to show the shadow of the gall-bladder filled with tetraiodophenolphthalein. The compound was given by mouth about 15 hours before the photograph was taken. Recently new compounds containing over 60 per cent. of iodine have come into use for displaying the gall-bladder and ducts. They can be given intravenously.

(By courtesy of R. C. Garry.)

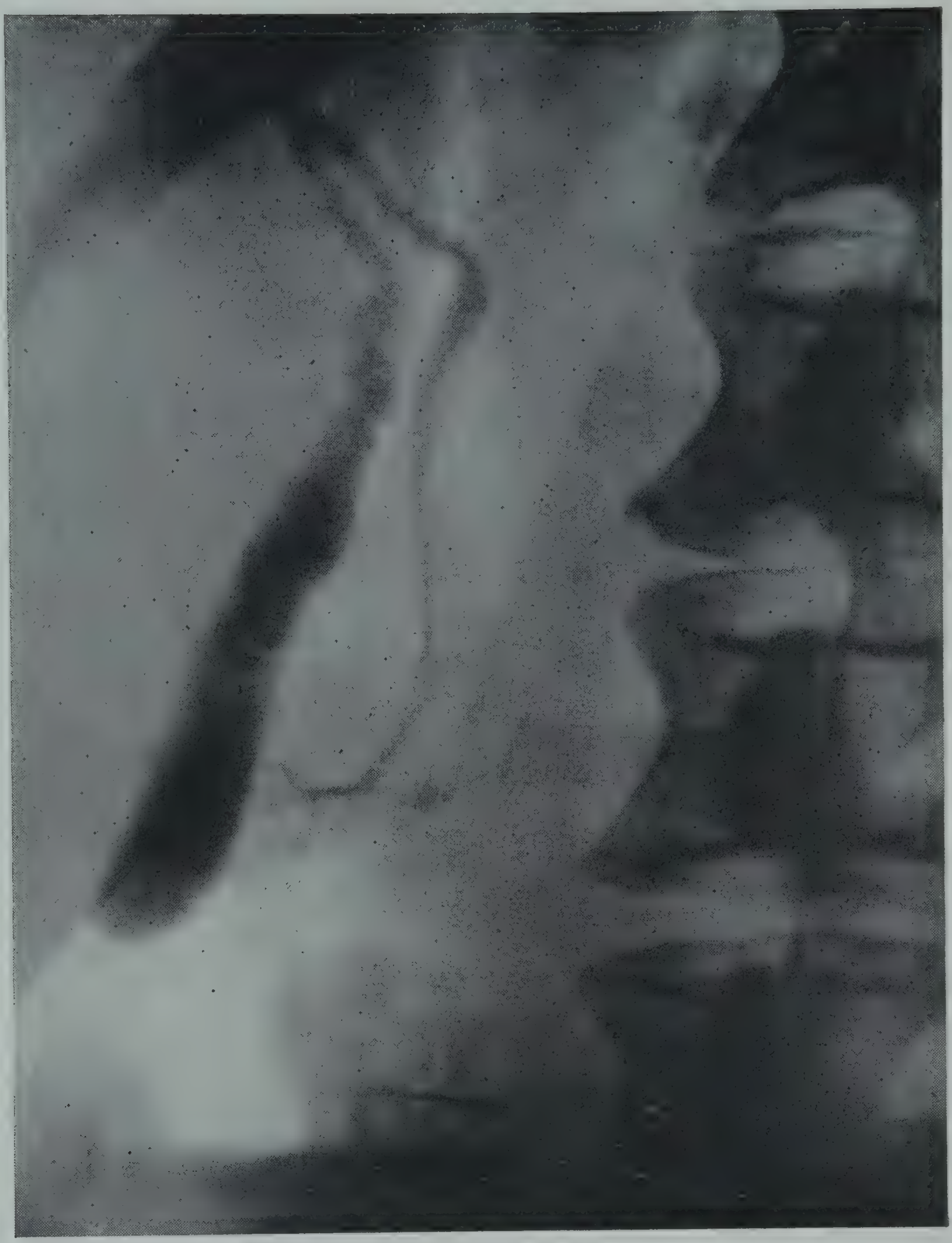


FIG. 17, 6

X-ray photograph of right subcostal region taken one hour after a fatty meal and sixteen hours after administration of an iodine compound. The gall-bladder has contracted in response to the meal; it is now one quarter of the size seen at fifteen hours. The cystic duct and the bile duct are both visible and the dye has passed into the second part of the duodenum where it is seen outlining the mucosal pattern. Such a complete demonstration of the bile duct is unusual. The normal gall-bladder is empty of dye four to six hours after a fatty meal.

(*By courtesy of C. Pickard.*)

been obtained in crystalline form is, like all enzymes, a protein. *Chymotrypsinogen*, which has also been isolated in crystalline form, is another enzyme precursor present in the pancreas. It is activated by trypsin to form the active enzyme *chymotrypsin*, which has been crystallized; it acts in very much the same way as trypsin, but in addition it possesses the power of clotting milk. Both trypsin and chymotrypsin are powerful protein-splitting enzymes with an optimum pH of about 8. They break down proteins and proteoses into peptides.

Although the reaction of the duodenum is usually on the acid side of neutrality (pH about 5) tryptic digestion proceeds much faster *in vitro* at alkaline reactions (pH 8 or 9). Trypsin, therefore, is not acting at its optimum pH in the duodenum. In alkaline solution trypsin is easily destroyed and even at the pH of the intestine it is slowly destroyed; as a result the intestinal contents from the lower parts of the intestine have little tryptic activity. It seems that the loss in activity at the pH of the intestine is offset by the gain in stability of the enzyme. The destruction of trypsin is inhibited by proteins and their breakdown products.

The pancreas also produces the enzyme *carboxypeptidase* which splits off from peptides the terminal amino-acids carrying the carboxyl group.

It is not until the food reaches the duodenum that the digestion of fats begins. The pancreatic lipase is a powerful enzyme which hydrolyses fat into a mixture of fatty acids and monoglycerides (p. 19). *Pancreatic amylase* has an action similar to that of salivary amylase (ptyalin), and like ptyalin requires a certain concentration of chloride ion for its activity. It is more powerful than ptyalin and can hydrolyse unboiled starch to maltose. Its pH optimum is 6·7 to 7·0. A certain amount of pancreatic amylase is absorbed into the blood and excreted in the urine which, therefore, possesses starch-splitting (diastatic) properties. In certain diseases of the pancreas the serum amylase activity and the diastatic power of the urine are enormously increased, providing an important diagnostic test.

The pancreatic juice also contains important *nucleases* which break down nucleic acids with the liberation of simple nucleotides (p. 63).

After pancreatectomy or complete ligation of the ducts in animals the digestion of fat and nitrogeneous substances is seriously impaired but the absorption of carbohydrate is scarcely affected. Similar disturbances are found in pancreatic disease in man.

The Control of Pancreatic Secretion

Stimulation of the parasympathetic vagal fibres to the pancreas produces after a brief latent period a small amount of pancreatic juice which is turbid and viscous and rich in enzymes. This secretion is quite independent of the chemical effect of the gastric hydrochloric acid described below, since it occurs even when the pylorus is ligated. Atropine abolishes this secretion and pilocarpine provokes it.

Stimulation of the sympathetic nerve supply derived from the cœliac plexus sometimes produces a small flow of pancreatic juice.

Bayliss and Starling gave the first clear proof in 1902 that a chemical mechanism is also involved in the secretion of pancreatic juice. They found that an acid extract of the duodenal mucous membrane injected, after neutralization, into the blood stream of animals caused a 'copious flow of pancreatic juice,' and they attributed this effect to the action of a hormone which they called *secretin*. This substance has been prepared in crystalline form and has been shown to be a protein of molecular weight about 5000. When the pure substance is injected intravenously the pancreas secretes profusely in less than a minute. The minimal amount of the pure substance needed to stimulate the pancreas is of the order of a few μg. The pancreatic juice produced by secretin is thin and watery, with a low enzyme content, but it is rich in alkali. Secretin produces this effect even when the nerve supply to the pancreas has been cut or when the pancreas has been transplanted. When given by the mouth secretin is ineffective as it is destroyed not by pepsin or trypsin but by a specific enzyme, secretinase.

A second chemical mechanism was described in 1943 by Harper and Raper who isolated from duodenal mucosa a factor called *pancreozymin* which can be separated chemically from secretin. On intravenous injection into animals it produces a secretion of pancreatic juice which is rich in enzymes. This response is independent of the nervous system, since it is unaffected by cutting the vagi or the splanchnic nerves, or by the administration of atropine. After injection of secretin the zymogen granules of the pancreas are little affected but after injection of pancreozymin the granules are greatly diminished (Fig. 17, 4).

Between meals there is in man a continuous but usually small secretion of pancreatic juice. A free flow of juice begins within one or two minutes after a meal and continues for about three hours. Initially the secretion is brought about reflexly, but it is continued when the acid gastric contents reach the duodenum and liberate secretin and pancreozymin. There are thus cephalic and intestinal phases of pancreatic secretion but no gastric phase. Pancreatic activity rapidly diminishes when the upper part of the intestine becomes empty again as the products of digestion pass on to the lower parts of the intestine which contain very little secretin or pancreozymin.

The volume and quality of the pancreatic juice vary according to the particular type of diet which the subject is accustomed to eat; a partial explanation for this is that the quantity of hydrochloric acid produced in the stomach and its rate of discharge into the duodenum are both affected by the character of the food. On a high carbohydrate diet the juice is rich in amylase and relatively poor in trypsin. On a high protein diet the juice is rich in trypsin and contains much less amylase. On a high fat diet the lipase content is not increased but the amylase is greatly reduced. The pH of the duodenal contents after a meat meal falls only to 4 or thereabouts, because the meat neutralizes to some extent the gastric contents, but it has been shown in dogs that this pH is low enough to make the pancreas secrete. Acid is undoubtedly the most effective agent for releasing secretin but it cannot be the only one since even in achlorhydria intestinal digestion may be unimpaired.

The best, but yet not entirely satisfactory, method so far described of obtaining duodenal juice in man is to pass down the œsophagus two separate tubes, one ending in the stomach and one ending in the duodenum. Continuous aspiration of both tubes prevents the stomach contents from mixing with the duodenal. The duodenal aspirate contains mainly pancreatic juice but also a certain amount of bile and intestinal juice.

Recently preparations of secretin and pancreozymin pure enough for use in man have been obtained. After secretin is injected into man the duodenal juice increases in volume, has a high bicarbonate content and a low chloride content, and there is also an increase in the output of all the enzymes. This injection can be followed in 30 minutes by an injection of pancreozymin. This produces a second increase in the volume of the duodenal juice, an increase of its bicarbonate content and an increase in the output of amylase, trypsin and lipase. In this case the duodenal juice is coloured with bile expelled from the gall-bladder which contracts under the influence of cholecystokinin which the extract contains, possibly as a contaminant. By the use of this method the normal response to standard doses of the two hormones in healthy persons has been measured but, since there is a wide scatter of results, it is difficult to say whether such a test will be useful diagnostically.

BILE

Bile is a slightly alkaline, viscous, golden-yellow or greenish fluid with a bitter taste. It is secreted continuously at a rate of 500 to 1000 ml. per twenty-four hours by the parenchymal cells of the liver and is subsequently concentrated in the gall-bladder from which it passes intermittently into the intestine in response to the arrival of food in the duodenum. Complete denervation of the liver leaves bile secretion unaltered; even the liver excised from an animal may continue to form bile.

The composition of human bile, as secreted by the liver and as obtained from the gall-bladder, is shown in Table 17, 7. Bile is a highly complex fluid and is not to be regarded simply as the secretion of a digestive gland. Its composition varies from time to time as certain of the biliary constituents represent material in the process of excretion from the body.

TABLE 17, 7

Composition of Human Bile (per cent.)

	Liver Bile	Gall-bladder Bile
Water	98·0	89·0
Total solids	2·0	11·0
Bile salts	0·7	6·0
Bile pigments	0·2	2·5
Cholesterol	0·06	0·4
Inorganic salts	0·7	0·8
pH	8·0 to 8·6	7·0 to 7·6

The bile salts are the sodium salts of glycocholic and taurocholic acids, which are formed in the liver by the conjugation of glycine and taurine (p. 355) with cholic acid or compounds related to it, such as lithocholic acid or deoxycholic acid (p. 27). In human bile glycocholic acid is the more abundant.

The bile salts have four main properties: (1) They react with many substances such as stearic acid, cholesterol and fat-soluble vitamins, which are not soluble in water, to form complexes termed choleic acids, which are water-soluble. This property is described as the *hydrotropic power* of the bile salts. The formation of the complexes is, therefore, of great importance in promoting the absorption of these insoluble substances. (2) By virtue of the ability of the bile salts to reduce surface tension the fat emulsion in the intestine is stabilized and the particles of fat are made smaller so that the surface area of fat exposed to the action of lipase is many times increased. (3) Lipases present in the small intestine are activated by bile salts. This effect is not due entirely to the hydrotropic action of the bile salts on fats, since it can be shown even when the substrate is a soluble ester. If bile is prevented from reaching the duodenum a quarter or more of the fat intake may be lost in the fæces. This suggests that the bile salts, as well as activating lipase, aid absorption of fatty materials. (4) Although bile salts are poured into the duodenum only small amounts are found in the fæces, since they are reabsorbed for the most part and returned to the liver by the portal vein to be re-excreted in the bile—the *enterohepatic circulation.* A small part of the bile salts is destroyed in the intestine by micro-organisms and a little passes out in the fæces, but this loss is made good by synthesis. The circulation of the bile salts is an important factor in promoting a continuous secretion of bile by the liver cells, for the bile salts themselves act as powerful *choleretics.* Bile salts given by the mouth or cholic acid given intravenously stimulate the formation of bile even when the liver is denervated. When the bile is drained away through a fistula, and not allowed to enter the duodenum, the enterohepatic circulation ceases and the flow of bile is greatly reduced. The hormone secretin, in addition to its effect on the pancreas, increases the rate of secretion of bile by the liver but it is a less potent stimulus than the bile salts. Secretin probably increases only the volume of bile without altering the amount of salts excreted.

Cholesterol is present in the bile no doubt as an excretory product, since the amount varies with the amount in the blood. It is kept in solution by bile salts as a choleic acid complex; hence any reduction in bile-salt concentration tends to throw cholesterol out of solution. Cholesterol deposited in solid form in the biliary passages constitutes one variety of gall-stone.

The *bile pigments* are to be discussed on p. 415. They are responsible for the colour of the bile which in man is golden-yellow owing to a preponderance of bilirubin. Although bile pigments are excretions with no digestive function, an animal completely deprived of its bile gradually becomes anæmic, its bones become decalcified and its health declines. This may be due to the loss of bile pigment since the animal's condition improves if it is fed with blood. The absorption of iron from the alimentary tract is facilitated by the presence of bile.

Bile may act as a vehicle for the excretion not only of cholesterol and bile pigments but of many other substances, including certain drugs. This fact is made use of in tests devised to assess the function of the liver. Thus the concentration in the blood at a standard time after the injection into the body of a dye, phenoltetrabromphthalein sodium sulphonate (bromsulphalein), known to be eliminated solely by the liver into the bile may serve as an index of hepatic function. The fact that certain radio-opaque iodine compounds, after intravenous injection or after absorption from the intestine, are excreted in the bile and concentrated in the gall-bladder permits the examination of this organ by means of radiography (Figs. 17, 5 and 17, 6).

The *inorganic salts* in bile include chlorides and bicarbonates, the latter being responsible for the slightly alkaline reaction of the bile. Small amounts of mucin and of nucleoprotein are also present.

The only enzyme present in bile is an alkaline phosphatase probably derived from liver cells but its role in the process of digestion in the gut is not yet clear. When the flow of bile into the intestine is impeded or prevented (obstructive jaundice) this phosphatase is found in the blood in increased concentration.

THE GALL-BLADDER

The gall-bladder has a wall consisting of a thin layer of muscle and elastic tissue lined by a layer of columnar epithelial cells, and can hold about 32 ml. as estimated from the shadow cast in the dye method (Fig. 17, 5). From the neck of the gall-bladder the *cystic duct* passes to join the common hepatic duct as it emerges from the liver. The *bile duct* so formed conveys the bile to meet the pancreatic secretion in the *ampulla of the bile duct* (Fig. 17, 1). All these ducts in man have a rather sparse supply of muscle fibres in their walls. As the duct goes through the duodenal wall it is surrounded by the sphincter of Oddi which can act independently of the surrounding duodenal musculature. Although its removal produces little serious inconvenience, the gall-bladder normally stores and concentrates bile and regulates its discharge into the duodenum. Table 17, 7, shows that there is a considerable difference in composition between the liver bile and the bile as it leaves the gall-bladder. During its stay in the gall-bladder the bile is concentrated about tenfold, mainly by the absorption of water and inorganic salts, and it becomes much darker in colour; at the same time its viscosity in increased by the addition of mucus secreted by the cells of the gall-bladder wall. The absorption of bicarbonate results in a fall of pH from 8·2, in bile leaving the liver, to about 7·0 in the gall-bladder. When the biliary tract becomes obstructed below the cystic duct the retained bile is progressively concentrated by the gall-bladder to an almost tarry consistency which favours the laying down of gall-stones. If the hepatic ducts are obstructed no such concentration occurs and the formation of gall-stones is less likely. When the gall-bladder has become functionless as the result of disease of its wall, or when the cystic duct is blocked by a stone, obstruction of the bile duct leads to an accumulation of the thin watery secretion from the walls of the ducts sometimes called *white bile*.

Movements of the Gall-bladder and Sphincter of Oddi.—The movements of the gall-bladder have been studied by Elman and McMaster in the unanæsthetized dog by cannulating the cystic duct, the common hepatic duct and the duodenal end of the bile duct as shown in Fig. 17, 8. In this way the three phenomena of secretion of bile, gall-bladder activity and the behaviour of the sphincter of Oddi, can be investigated separately or in combination as in the normal animal. The behaviour of the sphincter of Oddi can be tested by attaching a reservoir and a manometer filled with bile to the duodenal end of the bile duct and raising the pressure until bile flows through the sphincter into the duodenum. When tested four to twelve hours after a meal the pressure required is about 110 mm of bile but after a longer fast the sphincter does not open until the pressure is raised to 300 mm of bile. When the animal is shown food the sphincter gives way at a pressure of say 60 mm of bile but the resistance to flow increases somewhat during feeding.

If the hepatic duct is connected directly to a manometer so that no bile can escape, the pressure rises to a maximum value of 320 mm of bile in less than half an hour and then bile secretion ceases. If, however, the hepatic duct is connected to the gall-bladder while bile is prevented from flowing along the duct the pressure remains constant about 100 mm of bile for several hours. The gall-bladder by absorbing water must therefore be storing bile as fast as it is formed. Since the potential secretion pressure of the liver is much greater than 100 mm of bile there is no difficulty in understanding how the gall-bladder is kept supplied with bile. Again, since a pressure greater than 100 mm of bile is necessary to open the sphincter of Oddi between meals, the bile must be diverted to the gall-bladder and cannot enter the duodenum. These values have since been found to apply to man.

When food is eaten there is an immediate rise of pressure in the gall-bladder but the cannula in the hepatic duct does not show any increase in the flow of bile until after an hour or so, and even then there is no sudden gush (Fig. 17, 9). Since, at the moment of feeding, the resistance of the sphincter of Oddi is greatly reduced it is easy to understand that in the intact animal there is a gush of bile into the duodenum. This soon ceases and thereafter bile is discharged intermittently, only a small amount

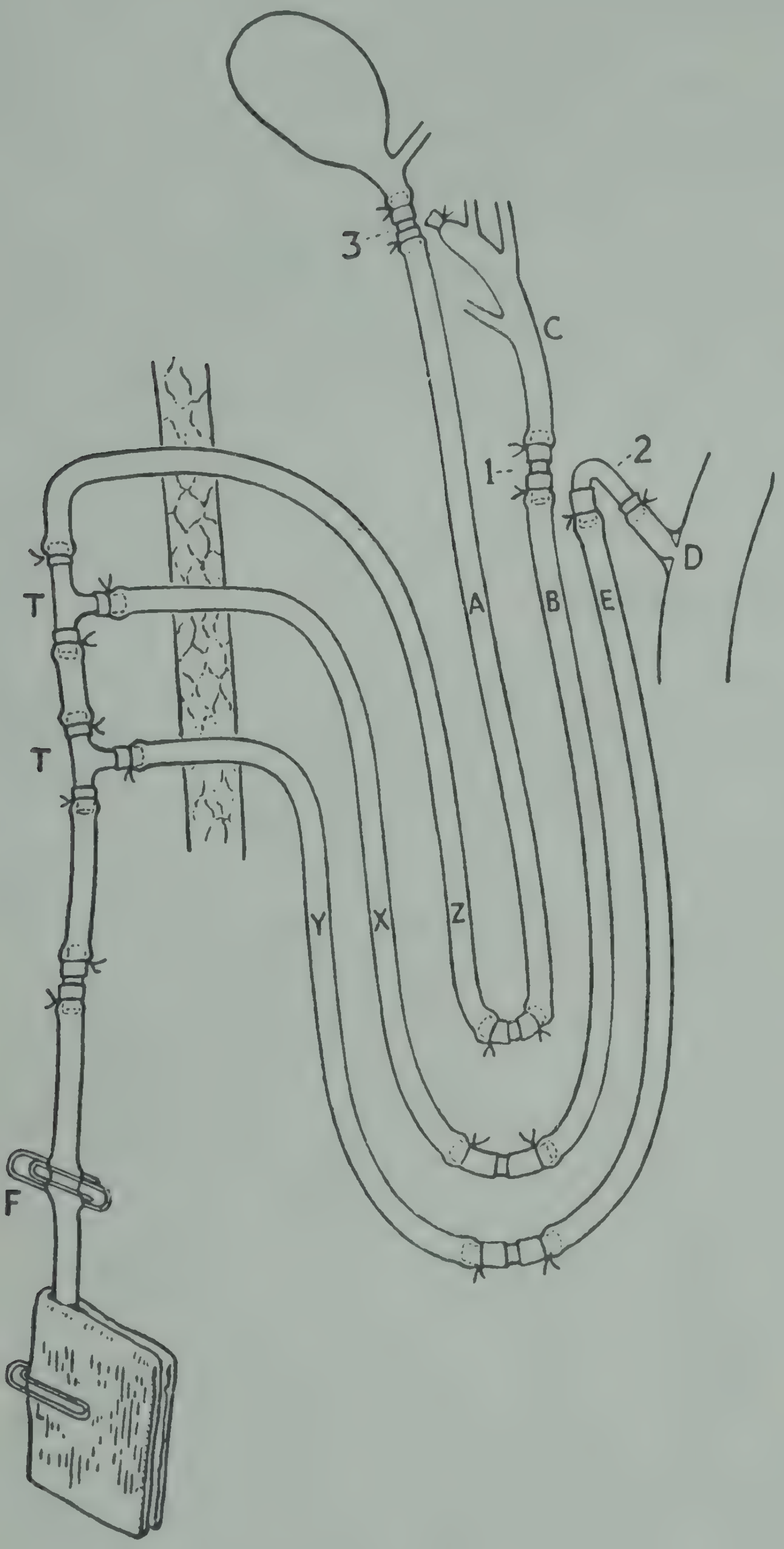


FIG. 17, 8

Plan of the 'triple intubation.' Besides the tubes XB and YE leading from and to the bile duct (C) which constitute the so-called 'altercursive intubation' there is a third tube ZA communicating with the gall-bladder. When the tube was clamped at F bile flow to the duodenum (D) went on as usual for weeks or months through the tubes T and T. There was merely a detour of the fluid to the outside (P. D. McMaster & R. Elman (1926), *J. exp. Med.* **44,** 184).

being ejected at any one time. It is possible that the actual ejection of bile is synchronous with the flow of acid chyme through the pylorus.

In man after double vagotomy there is no alteration of the rate of emptying of the gall-bladder but there is a relaxation of the tone of its walls so that the fasting volume is doubled. Emptying in man is not a reflex phenomenon but is almost certainly dependent entirely on the hormone cholecystokinin.

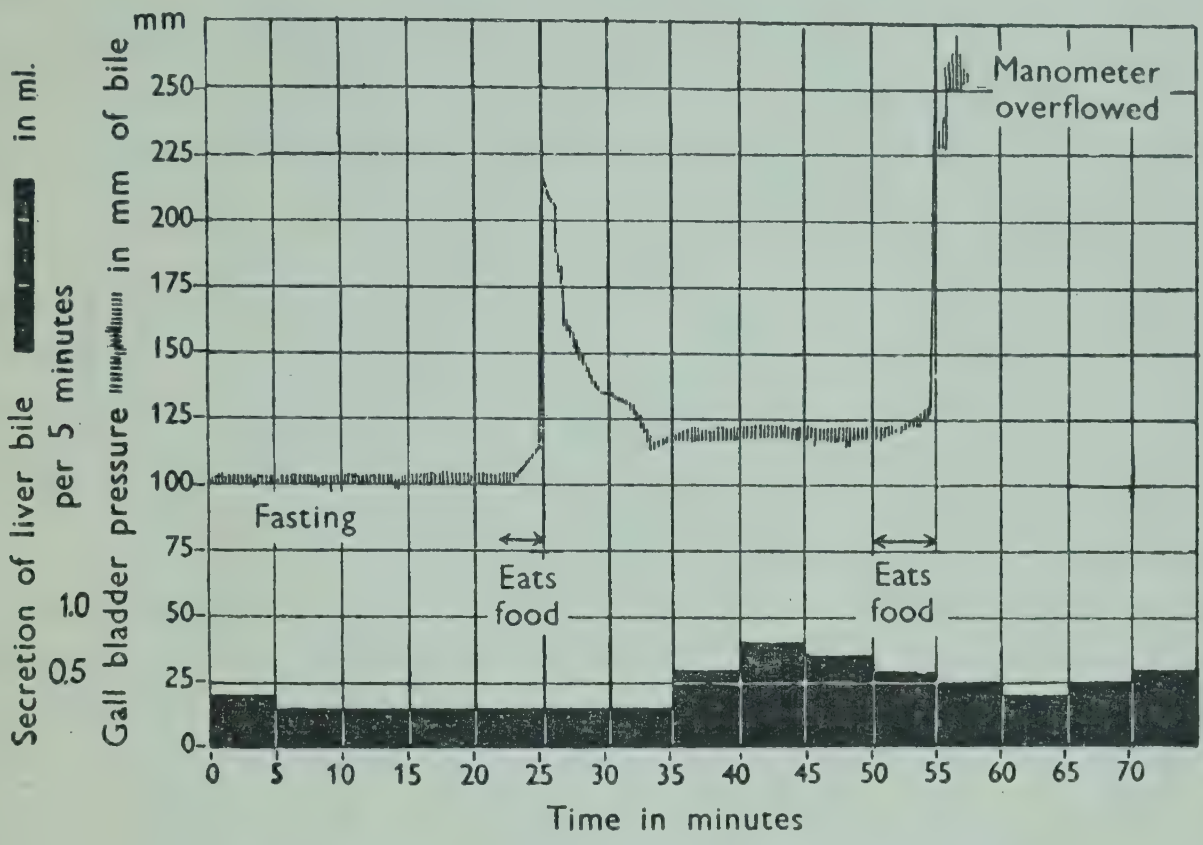


FIG. 17, 9

Pressure changes within the gall-bladder after the taking of food. The upper curve shows the sudden increases in gall-bladder pressure, immediately after eating. The lower columns show the amount of bile secreted by the liver in 5-minute periods (P. D. McMaster & R. Elman (1926), *J. exp. Med.* **44,** 185).

When acids or partly digested protein or fatty material, particularly egg-yolk, are placed in the duodenum the gall-bladder contracts even when it is denervated or transplanted into the neck. This is probably due to the liberation by the mucosa of a chemical substance into the blood stream. Extracts of duodenal mucosa have been prepared which on intravenous injection cause a contraction of the gall-bladder and a relaxation of the sphincter of Oddi. The name *cholecystokinin* has been proposed for the active principle, a protein related to secretin but quite distinct from it with a different isoelectric point, namely pH 5 to 5·5.

Strong solutions (25 to 33 per cent.) of magnesium sulphate taken by mouth cause a flow of bile into the duodenum. This may be due to a contraction of the gall-bladder or more probably to a reduction in the resistance to inflow into the duodenum. Both morphine and pilocarpine cause a spasm of the sphincter of Oddi in man. This is in conformity with the finding in the dog that stimulation of the vagus causes a spasm of this sphincter.

It is important to distinguish between *choleretics*, that is substances which increase the secretion of bile by the liver, and *cholagogues*, which bring about the contraction of the gall-bladder. The only substances in the former class are the bile salts and secretin; fatty foods, cholecystokinin and perhaps magnesium salts can be classed as cholagogues.

THE MOVEMENTS OF THE SMALL INTESTINE

The Duodenum.—The duodenum, the first portion of the small intestine, curves round the head of the pancreas. The first two inches (superior portion or duodenal bulb) go upwards from the pylorus and then the duodenum turns abruptly down for three or four inches (descending portion), when there is another abrupt turn to the left into the horizontal portion. The bile duct and the pancreatic duct usually enter the duodenum at the middle of the descending portion. The two ducts may enter separately or together, the terminal portions of both being surrounded by a sphincter muscle (muscle of Oddi, Fig. 17, 1).

The actual details of the movements of the duodenum are extremely variable, but the bulb fills more or less completely when the gastric contents pass through the pylorus. Immediately, or after a short delay, an annular wave of contraction, which may be initiated by distension, begins in the bulb and passes along the duodenum. On most occasions the wave dies out before reaching the jejunum.

Jejunum and Ileum.—In the small intestine the contents are thoroughly mixed by the movements described as *segmentation*. A length of intestine observed under X-rays is suddenly divided up by the contraction of the muscular wall into a number of approximately equal segments (Fig. 17, 15). A few seconds later a contraction occurs in the middle of each distended segment, and at the same time the contracted muscle between the segments relaxes. In this way a new set of segments is formed, to be replaced in a few seconds by the original segments. Obviously this kind of activity produces no onward movement of the intestinal contents, but it mixes them thoroughly with the digestive juices and brings them into contact with the mucosa so that absorption may take place.

The segmenting contractions are probably excited by the distension of the intestine produced by the arrival of a meal. Nerve paralysants such as cocaine and nicotine do not affect these movements and they can even be observed in strips of denervated muscle examined *in vitro*. They are not affected by incisions encircling the gut and carried down to the submucous coat which interfere with peristalsis. Since these rhythmic contractions must be dependent on the activities of the muscle fibres themselves they are said to be *myogenic*.

If strips of intestine are cut out and suspended in a bath of Locke's solution at 37° C and connected to a writing lever they show fairly regular rhythmic contractions and relaxations. If a piece of duodenum is used, the contractions may occur every five to ten seconds, but the rhythm is slower in samples taken from the jejunum and still slower in the ileum. There appears therefore to be a gradient of activity in the intestine, the proximal portions being more active than the distal. The movements are inhibited and the gut relaxes when adrenaline is added to the bath or when the nerves running along the blood vessels to the gut are stimulated electrically. Acetylcholine, on the other hand, produces a contraction of the intestine. The activity of the intestine in the intact animal or

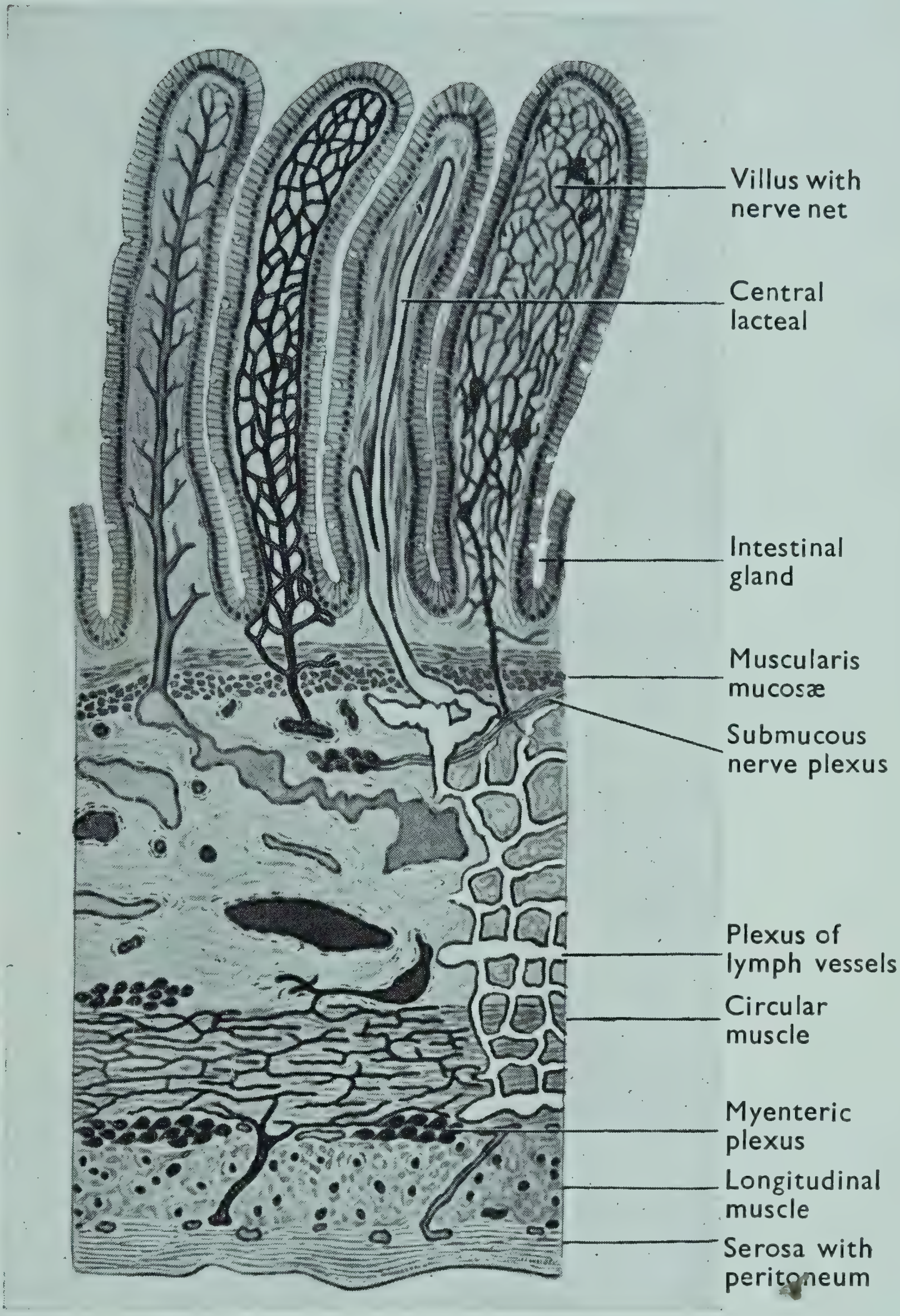

Villus with nerve net
Central lacteal
Intestinal gland
Muscularis mucosæ
Submucous nerve plexus
Plexus of lymph vessels
Circular muscle
Myenteric plexus
Longitudinal muscle
Serosa with peritoneum

Fig. 17, 10

Diagram of part of a cross-section through human small intestine. The villus on the extreme left shows the venous drainage and the adjacent villus shows the arterial supply.

(*After* Ramony Cajal.)

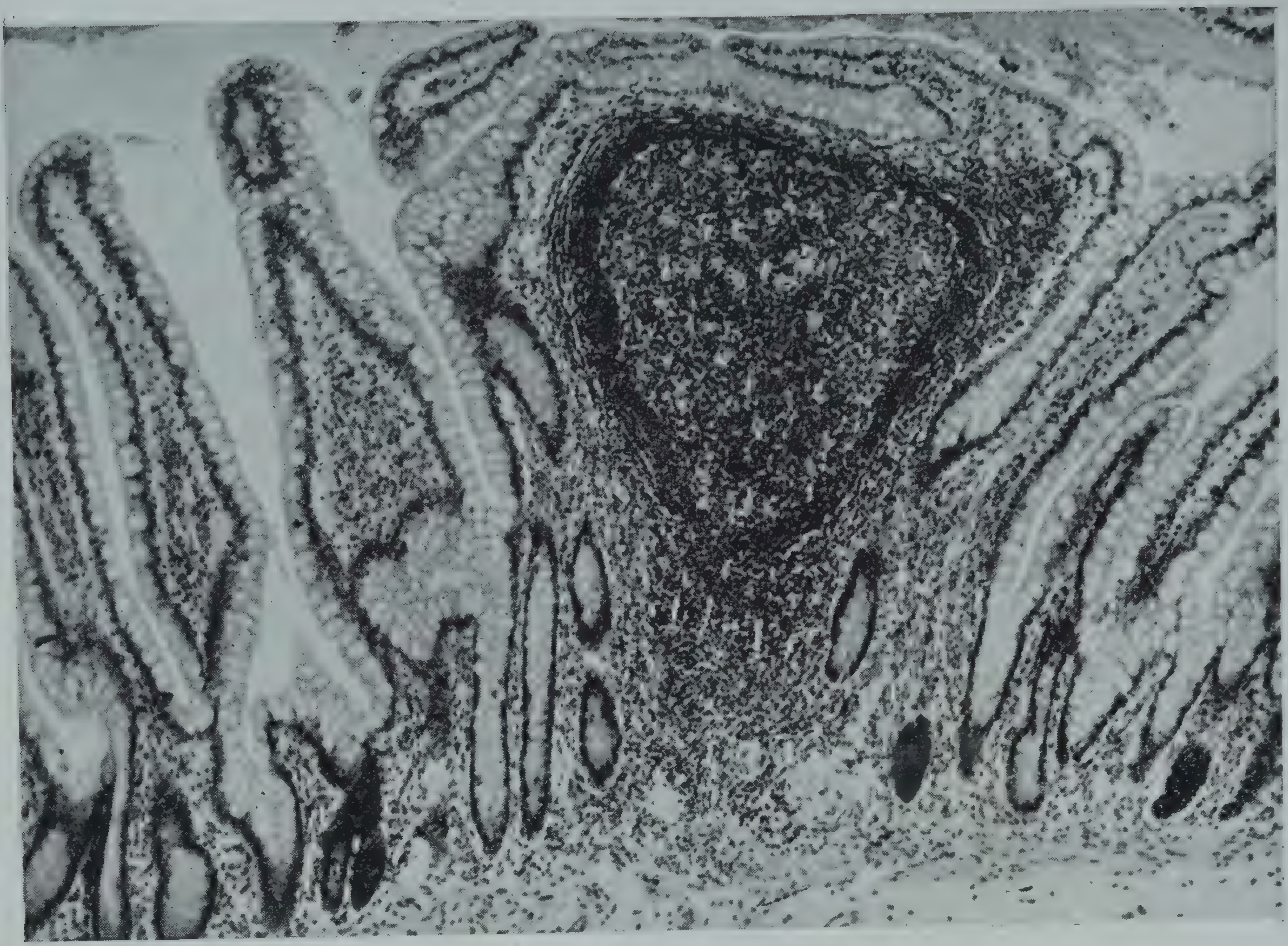


Fig. 17, 11

Section of human intestine (ileum) to show solitary lymph follicle with germinal centre. A few villi, pointing up into the lumen of the intestine, are seen on each side of the lymph follicle.

(By courtesy of A. C. Lendrum.)

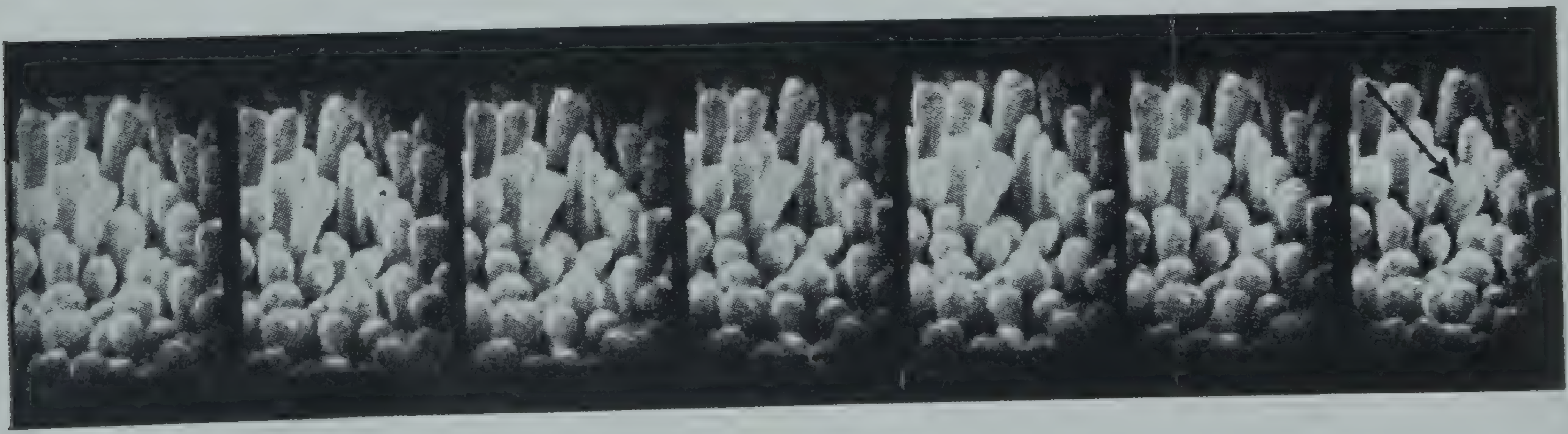


Fig. 17, 12

Movements of the intestinal villi in a dog. Every eighth frame of a cine-film is reproduced. Note the changes in the villus marked with an arrow (E. v. Kokas and G. v. Ludány (1930), *Pflüg. Arch. ges. Physiol.*, **225**, 421).

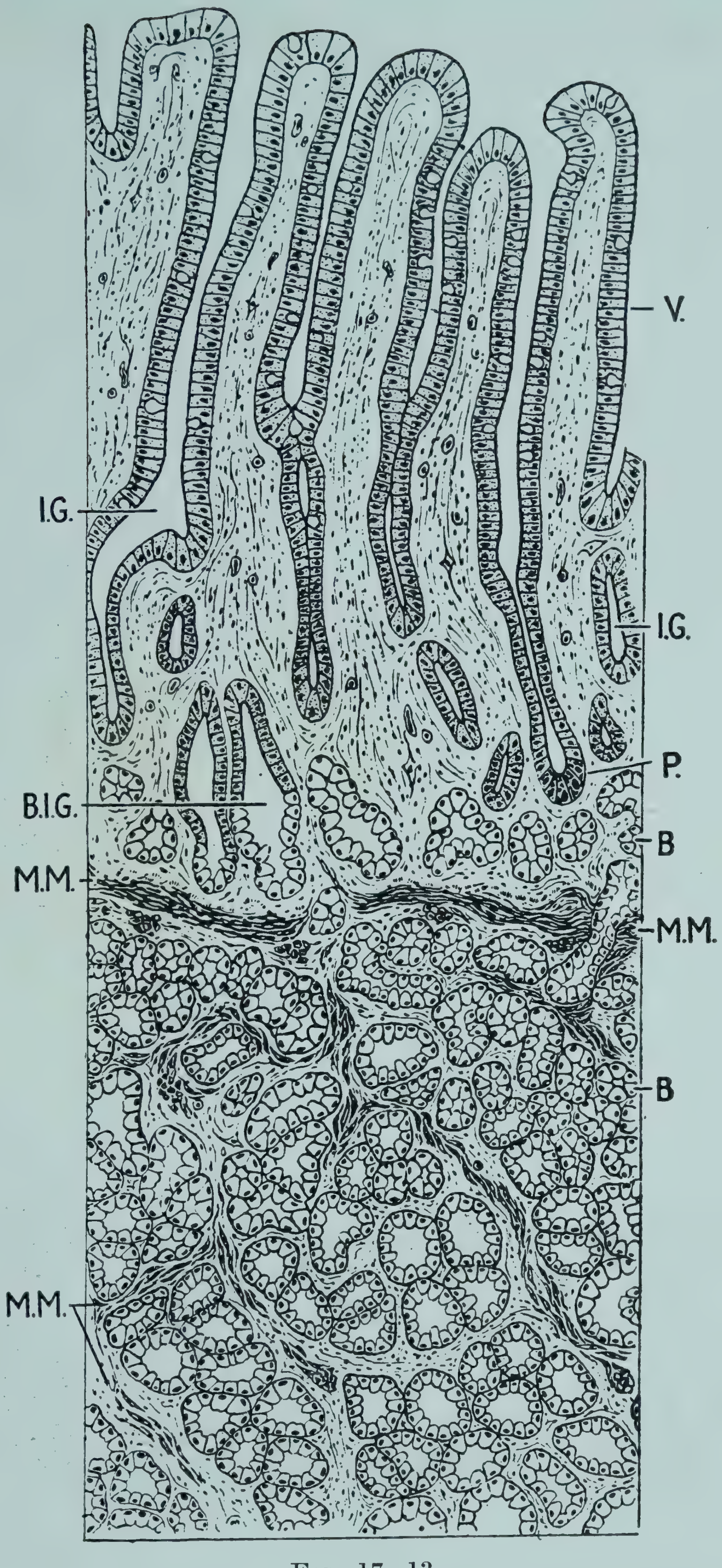


Fig. 17, 13

Section through the duodenum to show villi and the duodenal (Brunner's) glands. Most of these are deep to or surrounded by the muscularis mucosæ which is split up into many irregularly running strands.

V. = Villus.
I.G. = Intestinal glands.
P. = Paneth cell.
M.M. = Muscularis mucosæ.
B. = Duodenal (Brunner's) glands.
B.I.G. = Emptying of a Brunner gland into an intestinal gland.

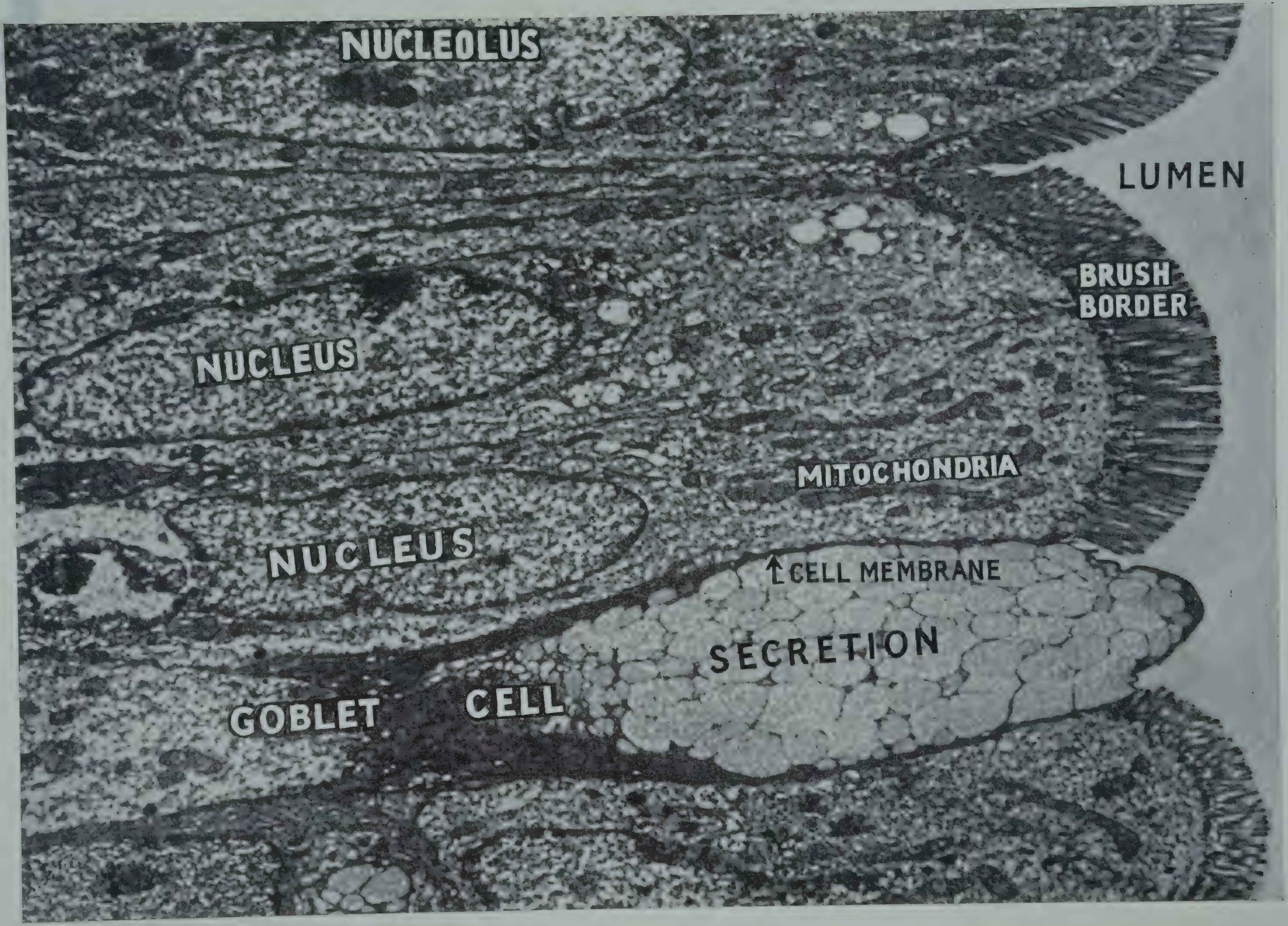

NUCLEOLUS
LUMEN
BRUSH BORDER
NUCLEUS
MITOCHONDRIA
NUCLEUS
CELL MEMBRANE
SECRETION
GOBLET
CELL

Fig. 17, 14

of the epithelium of the small intestine of a rat. Magnification 8000 times. Fixed in osmium tetroxide. Note human subject can be recorded by placing a balloon in the lumen and connecting it to a recording tambour. Stimulation of the vagus produces an initial inhibition followed by increased rhythmic activity and a rise of tonus. Stimulation of the splanchnic nerves *in vivo* has the same effects as *in vitro*, namely, relaxation of tone and inhibition of all movements but, contrary to expectations, after bilateral sympathectomy in man the motility of the small intestine is not changed. Finally slow alterations of tone have been seen in all parts of the gut and also pendular movements of loops of intestine; but their significance is unknown.

The onward movement of the intestinal contents is produced by *peristaltic waves*. These are contractions of the circular muscle of the gut behind a mass of intestinal contents preceded by a relaxation of the intestinal muscle. Bayliss and Starling about 1900 described peristalsis as a series of waves passing about 2 cm per minute along the gut, but Alvarez thinks this slow movement is unusual and that the true peristalsis is a much faster phenomenon—the waves travelling at 2 to 25 cm per second without any preceding relaxation. This peristaltic rush may pass along the whole of the small intestine.

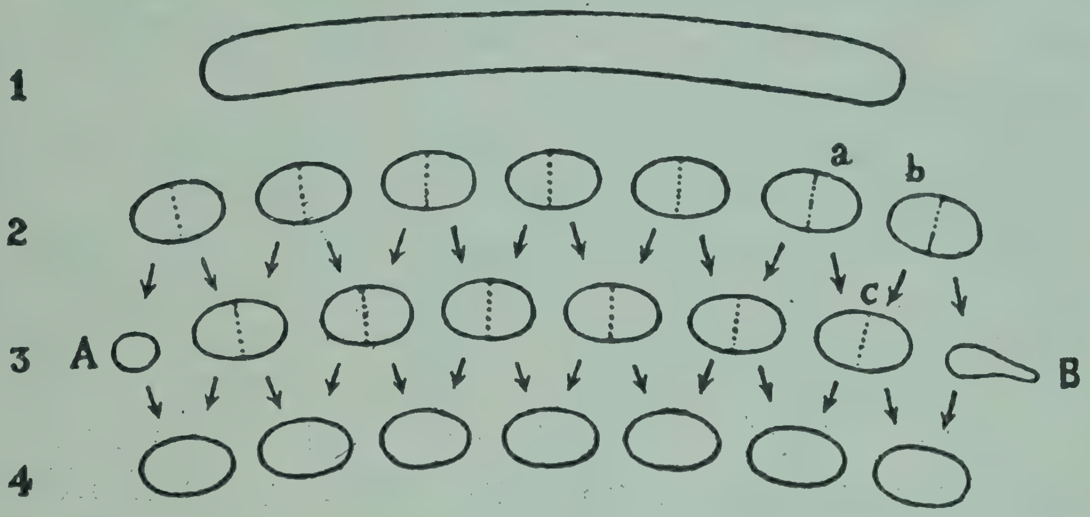


FIG. 17, 15

Diagram of the 'segmentation' movements of the intestines as observed by X-rays after administration of bismuth (Cannon, W. B. (1911), "The Mechanical Factors of Digestion." London: Arnold).

1. A continuous column, intestinal movements being absent.
2. The column broken up into segments.
3. Five seconds later, each segment divided into two, the halves joining the corresponding halves of adjacent segments.
4. Condition (2) repeated five seconds later.

If a balloon connected to a recording tambour is inserted into a loop of intestine, stimulation of the intestine by pinching it immediately craniad to the balloon is followed by an increase in volume of the balloon, that is by a relaxation of the wall of the intestine. Stimulation just caudad to the balloon is followed by a contraction of the muscular coat around the balloon. This *myenteric reflex* is dependent on the integrity of the myenteric plexus; it remains after section of the nerves going to the gut, but disappears if the myenteric plexus is put out of action by nicotine or cocaine. This local reflex explains how a peristaltic wave is carried along the intestine. A bolus in the intestine causes increased muscular activity on the oral side of the bolus. What appears to be a relaxation on the aboral side may be due to a contraction of the longitudinal muscle or simply to diminished tonus. The intestinal contents pass in the aboral direction since the contraction behind a bolus lasts so long that there is little chance of regurgitation. [Within a minute or so of beginning a meal a series of peristaltic waves passes from the duodenum along the intestine and carries the intestinal contents onwards through the ileocolic sphincter, which relaxes when the wave reaches it, into the large intestine. The intestine has some kind of peristaltic orientation since antiperistaltic movements have not been observed. Thus if a loop is cut in two places, then reversed and sewn up in continuity once more, the intestinal contents are held up.] Peristaltic waves started up by stimuli craniad to the reversed loop come to an end when the reversed portion of the gut is reached, presumably because they cannot be transmitted along the gut in the reverse direction. A partial obstruction occurs with dilatation of the gut on the cranial side of the reversed loop.

Distension of the intestine within physiological limits, as just described, stimulates the intrinsic mechanism to produce peristaltic movements but if the internal distending pressure is too high the movements of the entire small intestine may be inhibited. The afferent and efferent paths of this inhibitory reflex are in the thoracolumbar sympathetic system. Pathological distension of the gall-bladder, bile ducts, ureter or urinary bladder, may similarly cause inhibition of gastric and intestinal movements.

Antiperistaltic movements have been described in the duodenum and ileum, but if they occur they are quite feeble.

The *ileocolic sphincter* is formed, like the pyloric sphincter, by a thickening of the circular muscle fibres where the ileum enters the colon. Normally it is contracted so that the intestinal contents cannot escape too quickly into the cæcum. This allows adequate time for intestinal digestion and absorption. On arrival of a peristaltic wave the sphincter opens and quickly closes again preventing regurgitation from the cæcum. Stimulation of the splanchnic nerves causes the ileum to relax and the sphincter to contract. The vagus nerve, however, has no effect on the sphincter. No definite inhibitory outflow to this sphincter has been found. [The products of digestion of a meal reach the ileocolic sphincter in three or four hours and the last of the meal leaves the ileum some eight or nine hours after ingestion. The length of the intestine (mouth to anus) after death is usually given as 7 metres but it may be as little as 4 or as much as 8 metres. A narrow rubber tube 2·5 metres long can be passed from mouth to anus in a living man; the great lengthening which occurs after death must be due to loss of tone.]

There is much evidence, both clinical and experimental to show that intestinal motility is increased by a meal. Meals of different composition traverse the small intestine at different rates but digestion and absorption are complete when the residue reaches the end of the ileum. It could be assumed that the increased activity of the intestine is in some way co-ordinated with the requirements of digestion and absorption. The clue to this seems to be provided by Gregory's recent experiments in which he perfused Thiry-Vella loops of jejunum (Fig. 17, 18) at constant pressure with 0·9 per cent. sodium chloride and observed the rate of flow of the fluid through them. When a meal was fed there was within a few seconds an increase in tone and motility of the loop with a retardation in the rate of flow through it. The same result was produced by sham feeding, by distension of the stomach or by placing a small piece of food in the stomach. The effect is abolished by denervation of the loop. It may be that this reflex effect, in combination with the enterogastric reflex (p. 235) which restrains gastric activity, controls the rate of passage of the chyme along the intestine.

The Intestinal Villi.—The surface of the mucous membrane of the small intestine shows an enormous number of finger-like processes (*villi*), some 20 to 40 per sq. mm, which can scarcely be distinguished by the naked eye. Each villus (Fig. 17, 10) is covered by columnar epithelium and contains an arteriole and venule with their communicating capillary plexus and a blind ended lymphatic (or *lacteal*). The venules ultimately lead into veins of the portal system and the lacteals to the thoracic duct. The arrangement is such that blood passing through the minute vessels of the small intestine is brought into close proximity with the fluid material inside the intestine over an enormous area which, on the assumption that the free surface of the epithelial cells is smooth, is estimated to be 10 square metres. The electron microscope has shown, however, that the striated free border of the columnar epithelium is composed of minute processes about 0·6 μ long and 0·08 μ in diameter passing into the intestinal lumen (Fig. 17, 14). These processes increase the absorptive area of the epithelium at least thirty times.

The arteriole to each villus is connected directly to the corresponding venule by an arteriovenous anastomosis, which is open when there is no absorption but is closed during absorption, so that the blood may reach the capillaries in the apex of the villus.

In the upper part of the small intestine (jejunum) the villi are particularly large and the area of the mucosa is further increased by circular folds. The stroma of the mucous coat contains large numbers of lymphocytes which may be grouped together as solitary lymph follicles. These are particularly well developed in the ileum as *aggregated lymphatic nodules* or Peyer's patches (Fig. 17, 11).

The Movements of the Villi.—During digestion and absorption the villi are in constant movement. They have been observed under a binocular microscope through a hole in the antimesenteric area of the intestine. In the dog the villi shorten fairly quickly at irregular intervals and then relax slowly, each villus contracting about six times per minute quite independently of its neighbours (Fig. 17, 12). The submucous nerve plexus seems to control this activity, since mechanical stimulation or the application of eserine or amino-acids all produce movements of the villi if the submucous plexus is intact but not if it is stripped off. Local application of acetylcholine to the mucosa or close arterial injection of acetylcholine in minute amount increases the activity of the villi. Stimulation of the vagus causes a temporary increase of movements followed by inactivity; stimulation of the sympathetic causes the villi to become short, anæmic and motionless. It has been claimed that a specific hormone —*villikinin*—liberated by the acid from the stomach, acting on the duodenal mucosa, stimulates the movements. In cross-circulation experiments where the blood vessels of two dogs are joined up the application of hydrochloric acid to the duodenum of one dog causes movements of the intestinal villi in the other animal. Muscle fibres derived from the muscularis mucosæ pass into the villus to be attached to the lacteal and to the basement membrane just under the epithelium. It is reasonable to suppose that contraction of these fibres forces lymph into the deeper lymphatic vessels of the submucosa, which possess valves to prevent the back flow of lymph. Thus lymph may be pumped along the lacteals by the movements of the villi. In a fasting animal the movements are in abeyance and the villi are flattened so that they have an elliptical cross-section. When water or Locke's solution is applied to them movements immediately begin again and the villi swell and resume their rounded shape.

THE INTESTINAL JUICE (SUCCUS ENTERICUS)

Two distinct types of gland are found in the small intestine: (1) The *duodenal glands* (glands of Brunner), found only in the submucosa of the duodenum (Fig. 17, 13), have an alkaline secretion which contains mucin and a weak proteolytic enzyme. (2) The *intestinal glands* or crypts of Lieberkühn (Fig. 17, 10, and 17, 11), present throughout the small intestine, are simple tubular glands which open into the spaces between neighbouring villi. They possess a number of different types of cell which secrete mucus and a large number of enzymes. The *intestinal juice* or *succus entericus*, produced largely in the intestinal glands, has a pH of about 7·6 and contains some 1·6 per cent. of solid matter, of which about 1 per cent. is inorganic (Table 17, 16). The organic constituents include a great variety

Table 17, 16

Composition of Intestinal Juice (succus entericus) (pH 7·0 to 8·5)	
Organic constituents (0·6 per cent.)	'Erepsin' containing aminopeptidase and dipeptidases Amylase Maltase Lactase Sucrase Lipase Esterases Nucleases Nucleotidase Phosphatase Enterokinase Mucin (Some of these enzymes are intracellular enzymes released from mucosal debris)
Inorganic constituents (1·0 per cent.)	Na^+ Ca^{++} Mg^{++} Cl^- HCO_3^-

of enzymes which complete the digestive process although many of these enzymes are not true secretory products but are really intracellular enzymes liberated from cast-off mucosal cells.

1. '*Erepsin*' is the name given to a mixture of enzymes including amino-peptidases which remove from polypeptides the terminal amino-acid carrying the free amino group, and dipeptidases which split dipeptides to amino-acids.
2. *Amylase* brings about the digestion of starch to maltose which in its turn is broken down to glucose under the influence of *maltase*. At the same time the disaccharides, lactose and sucrose, are hydrolysed by *lactase* and *sucrase* respectively to their constituent monosaccharides, namely, galactose and glucose from lactose, fructose and glucose from sucrose.
3. *Lipase* is responsible for the breakdown of fat.
4. Nucleic acids are hydrolysed by *nucleases* to yield nucleotides. These in their turn are broken down, first by *nucleotidases* and then by the *nucleosidases*, to yield the constituent purine and pyrimidine bases. Nucleosidases, like phosphatase, probably occur mainly inside the cells of the mucosa.
5. *Phosphatase*, present in relatively high concentration in the intestinal villi, may play an important part in the absorption of fat (p. 270) and carbohydrate (p. 269). It is probable that it also is essentially an intracellular enzyme.
6. *Enterokinase*, already mentioned on p. 252, activates trypsinogen.

The Digestive Changes in the Small Intestine.—The *chyme* as it enters the duodenum from the stomach consists of a mixture of coarsely emulsified fat, protein (together with protein derivatives produced by the action of pepsin) and carbohydrates, including starch, a large proportion of which has escaped hydrolysis by ptyalin. The acid chyme is neutralized to some extent by the alkaline pancreatic juice and by the bile and the intestinal juice, but the reaction of the intestinal contents never becomes strongly alkaline and indeed tends to remain slightly acid. It should be understood that, while digestion is normally begun in the mouth and stomach, the enzymes secreted into the duodenum by the pancreas and the intestinal wall are capable, together with the bile, of initiating and carrying through the whole digestive process.

The *proteins* in the chyme are attacked first by the pancreatic endopeptidases *trypsin* and *chymotrypsin* which split the protein molecules into smaller fragments of the polypeptide type. These in turn are broken down by the exopeptidases, namely the *aminopeptidases* and *carboxypeptidases* (p. 253) which split off the terminal amino-acids, and the process of digestion to amino-acids is completed by the *dipeptidases* of the intestinal juice.

Carbohydrate enters the duodenum partly in the form of disaccharides such as maltose, produced by the action of ptyalin, or sucrose and lactose ingested with the food. In the intestine starch is hydrolysed to maltose under the influence of an amylase derived mainly from the pancreas, but to a smaller extent from the intestinal juice, and maltose, lactose and sucrose are hydrolysed to monosaccharides by the appropriate enzymes, maltase, lactase and sucrase. The final products of the digestion of carbohydrate are, therefore, the simple monosaccharides — glucose, galactose and fructose.

Cellulose is not digested in the human alimentary canal which lacks the appropriate enzymes. In the rumen of the ox and goat, however, cellulose is broken down by a variety of micro-organisms with the production of simple fatty acids such as acetic, propionic and butyric. The horse has an enormous large intestine in which bacterial digestion of cellulose occurs.

TABLE 17, 17

Digestive Enzymes

Anatomical Site	Secretion	Enzyme	Optimum pH	Substrate	Products of Enzyme Action
Mouth	Saliva	Ptyalin	6·7	Starch	Maltose
Stomach	Gastric juice	Pepsin	1·2	Proteins	Higher polypeptides
		Rennin (infants)	5 to 6	Caseinogen	Casein
Duodenum	Pancreatic juice	Trypsin and chymotrypsin	8	Proteins	Polypeptides
		Carboxy- and amino-peptidases	—	Polypeptides	Peptides and amino-acids
		Lipase	8	Fats	Fatty acids and monoglycerides
		Amylase	6·7 to 7·0	Starch	Maltose
		Nucleases	—	Nucleic acids	Nucleotides
Small intestine	Succus entericus	'Erepsin' (peptidases)	8	Peptides	Amino-acids
		Lipase	8	Fats	Fatty acids and monoglycerides
		Maltase	—	Maltose	Glucose
		Sucrase	6·2	Sucrose	Glucose and fructose
		Lactase	—	Lactose	Glucose and galactose
		Nucleotidase	—	Nucleotides	Nucleosides and phosphoric acid

Fats are attacked by lipase of the pancreatic and intestinal secretions but, as discussed later (p. 270), the extent to which they are hydrolysed is a matter of some doubt. In the upper part of the small intestine, partly under the influence of the bile salts, ingested fat is converted into a fine stable emulsion of particles less than 0·5 μ in diameter. Although the bile salts can bring about a great reduction in surface tension, they produce only relatively crude emulsions. The fine emulsion characteristic of the upper part of the small intestine was formerly attributed to the action of soaps formed in the intestine from the fatty acids liberated by digestion. Soaps, however, though excellent emulsifying agents, bring about emulsification only at pH values above 7·5. Since the pH of the intestinal contents is nearer 6·5 it is clear that soaps cannot play any important part in the emulsification of fat. Frazer has recently shown that a mixture of fatty acid, bile salts and lower glycerides is the only effective system present in the intestinal lumen capable of producing fine and stable emulsification of fat under the conditions that normally prevail in the intestine. He has also found that adequate quantities of monoglycerides are in fact produced by pancreatic digestion of fats.

Nucleic acids are broken down by nucleases, nucleotidases and nucleosidases to a mixture of purine and pyrimidine bases, phosphoric acid and pentose sugars.

A list of the digestive enzymes of the alimentary canal is given in Table 17, 17.

Control of the Secretion of the Small Intestine.—The secretory processes can be studied in animals by means of a Thiry-Vella fistula (Fig. 17, 18), which is made by cutting the intestine in two places. The separated loop is brought up to the abdominal wall and the continuity of the gut is restored as in Fig. 17, 18. There is little secretion from the fistula between meals and administration of food, if the loop is empty, produces little or no flow of juice. Mechanical stimulation of the loop, by the introduction of a rubber tube, for example, excites a flow of juice whether the animal is fed or fasting. Crude extracts of intestinal mucosa on injection produce a flow of juice but pure secretin does not do so. Claims have been made that a specific hormone is responsible for producing a secretion of succus entericus. It is probable that the flow of intestinal juice in the intact animal is initiated by the mechanical contact of the intestinal contents and the effect is reinforced by the products of digestion. The splanchnic nerves may exert an inhibitory effect on the secretion of intestinal juice, since a denervated loop at first secretes more juice than an innervated loop, but this may be due to the vasodilatation following on splanchnic section. Vagal stimulation or the injection of vagomimetic drugs (acetylcholine, pilocarpine) causes a secretion of succus entericus.

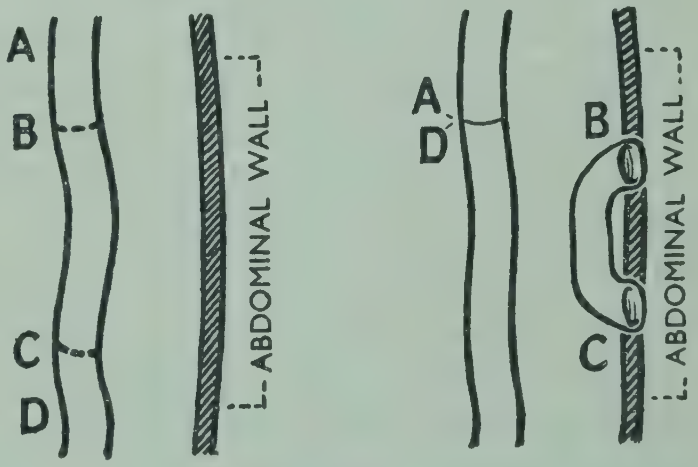


Fig. 17, 18

Thiry-Vella fistula.

An isolated loop of intestine is made by cutting the gut at two places, B and C, and then restoring the continuity at AD. The loop BC is then made to open on the abdominal wall.

ABSORPTION

The absorptive power of the intestine varies greatly in different regions. In the mouth and œsophagus no appreciable absorption of foodstuffs occurs although certain drugs, for example morphine and steroid hormones, are absorbed through the oral mucous membrane. Absorption through the gastric mucosa is very limited but small amounts of water, simple salts and glucose, as well as alcohol, may be absorbed. In the colon absorption is confined to water and water-soluble materials of low molecular weight such as glucose and inorganic salts. It is impossible to maintain metabolic equilibrium by introducing into the colon fluids containing protein and fat since they are not absorbed by the mucous membrane. Small portions of intestine can be excised and, provided that the continuity is restored, no permanent harm is done but removal of more than 50 per cent. is usually fatal because absorption is then deficient. After a large resection, especially of the ileum, there is a considerable loss of fat in the fæces and a small loss of nitrogen; absorption of carbohydrate is not affected. If an intestinal cul-de-sac is created in a patient in the course of an operation, perhaps to relieve obstruction, he is liable some time later to develop steatorrhœa and deficiencies of vitamins, especially B_{12}. The cause is not known but it is likely that there is a disturbance of the intestinal flora and an alteration in the composition of the intestinal juice.

Absorption occurs most actively in the upper part of the small intestine where the structure of the mucous membrane is specially adapted for the purpose (p. 263). Since the structure responsible for the absorption of material is the *villus* it will be evident that substances which are being absorbed must pass through (1) the outer or free border of the epithelial cell, (2) the body of the cell itself, and (3) the inner border of the cell before they reach the venule or lacteal. If the absorption of soluble substances from the gut is essentially a process of diffusion, it will obviously be facilitated by the presence of a favourable diffusion gradient. In other words, the greater the concentration of solute in the intestinal contents relative to that in the epithelial cells of the villus, or in the blood, the more readily will diffusion occur. However, in actual fact, the process of absorption is not so simple as this. Since after digestion the intestinal contents consist essentially of substances of low molecular weight they have a high osmotic pressure which tends to cause water to pass from the blood into the intestinal lumen; on the other hand, the colloid osmotic pressure of the plasma proteins tends to attract water from the lumen into the blood. Further, the hydrostatic pressure of the intestinal contents tends to force water out of the intestine while the hydrostatic pressure in the blood vessels favours the reverse process, that is the passage of water from the blood into the gut. The influence of these various forces may be seen in the case of the absorption of some simple sugars. Glucose is absorbed quickly, probably by a process of *selective absorption* (see next page). From an isotonic solution of glucose (5·4 per cent.) in the gut, water and glucose pass together into the blood, and, since the concentration of glucose in the gut is greater than that in the blood, this process occurs readily and absorption of both solvent and solute is rapid. If the glucose solution in the gut is hypertonic, then glucose passes readily into the blood, while water passes back into the lumen to dilute the intestinal contents until they are isotonic with blood.

Water and glucose are then absorbed together. If the glucose in the lumen is in hypotonic solution, water is absorbed rapidly until the solution becomes isotonic, when water and glucose are absorbed together. The absorption of sodium chloride takes place in a similar way, that is, from an isotonic solution.

This process may be contrasted with the absorption of *arabinose*, a sugar which is absorbed much more slowly, probably by a process of simple diffusion. From a solution of arabinose which is isotonic or hypertonic, arabinose slowly diffuses into the blood. At the same time water and sodium chloride pass back into the gut until the intestinal contents become isotonic, when all three are absorbed together. On the other hand, if the arabinose solution is hypotonic, water is absorbed readily at first, and then the slow absorption of arabinose, accompanied by the passage of sodium chloride back into the gut, occurs as before.

The absorption of substances from the gut is, however, not a purely mechanical process, since it involves considerable energy exchange. This is, moreover, not the same for all materials being absorbed. Again, the intestinal epithelium is able in many cases to exercise selective powers of absorption. Although sodium chloride is absorbed sodium sulphate is not, even in isotonic solution. Iron is absorbed only when there is a definite need for it in the body (p. 98) and, as is shown below, certain sugars are absorbed more rapidly than would be the case if the process consisted of simple diffusion.

The Absorption of Minerals.—The absorption of the important minerals such as calcium and iron is discussed in Chapter 9 (pp. 97 to 103).

The Absorption of Protein and Amino-acids.—During the process of digestion most proteins are broken down completely into the constituent amino-acids which pass readily through the intestinal wall into the portal blood and are carried to the liver. This is shown by the rise in amino-acid N in the portal blood during digestion. The absorption is selective as shown by the fact that L-amino-acids disappear from the gut more quickly and appear in the portal blood more quickly than do the corresponding D-amino-acids. Simple peptides may be absorbed in small amounts and in special cases even intact protein molecules. The latter process, however, occurs only in certain individuals in rather special circumstances. If a foreign (that is non-human) protein is absorbed unchanged into the blood antibodies are formed. When the concentration of these antibodies has been raised sufficiently by the repeated absorption of small doses of the specific protein (antigen), the reaction between the antibodies and a further dose of the antigen may produce asthma or skin rashes, and the patient is then said to be sensitive (or hypersensitive) to the protein.

The Absorption of Carbohydrates.—We have already seen that the main products of the digestion of carbohydrates are the simple monosaccharides. It might be expected that all hexoses would be absorbed at the same rate since they all have the same molecular weight. However, it can easily be demonstrated that glucose and galactose are absorbed more rapidly than fructose or mannose. The rate of absorption for most pentose sugars is nearly the same and is much lower than that for glucose or galactose; the pentoses might have been expected to be absorbed more quickly as they have a lower molecular weight than the hexoses. It is known that those sugars which are rapidly absorbed, that is glucose and galactose, can readily be phosphorylated in the body, and it is usually assumed that such sugars are absorbed by a process of selective absorption which involves phosphorylation within the cells of the intestinal mucosa. Sugars, such as mannose and the pentoses, which cannot be phosphorylated in the body, are absorbed by the slower process of simple diffusion. The effect of phosphorylation is to reduce the effective concentration of free sugar within the cell as fast as it enters, with the result that a very steep diffusion gradient for glucose and galactose is attained and these sugars are, therefore, absorbed very rapidly. The opposite process of dephosphorylation is supposed to occur as the sugars leave the intestinal mucosa and enter the portal blood, where they are present in the free state and not as sugar phosphates. Phlorrhizin, which inhibits phosphorylation processes, has been shown to reduce intestinal absorption of glucose and galactose.

The Absorption of Fat.—It was formerly thought that the action of lipase in the intestine was to break down fats completely to fatty acids and glycerol which were absorbed separately into the intestinal cells, the fatty acids being made soluble by the hydrotropic action of bile salts. Here they were recombined to yield first phospholipids and then neutral fats, which passed to the lacteals for transport via the thoracic duct to the systemic blood. Frazer has recently shown that complete hydrolysis of glycerides is not a step essential to their absorption. *In vitro* studies on pancreatic lipolysis revealed fatty acids and di- and mono-glycerides as end products, with extremely slow formation of free glycerol. The formation of these lower glycerides has also been shown in the intestinal lumen. Frazer has followed the course of fat absorption in rats by feeding fat stained with a suitable dye so that its movement could be traced. When neutral fat (olive oil) was fed, the fat was finely emulsified in the intestinal lumen, the lacteals became milky and a pronounced systemic lipæmia was observed. The stained fat was deposited mainly in the fat depots, little being found in the liver. On the other hand, after the feeding of stained fatty acid (oleic acid) and glycerol the picture was quite different. The fatty material was not emulsified in the intestinal lumen, the lacteals were not milky, no systemic hyperlipæmia occurred and the fat was deposited in the liver rather than in the depots. This evidence suggested that fat was not necessarily absorbed in the form of fatty acid. To verify this conclusion Frazer fed neutral fat plus an active preparation of lipase to rats, and found that the results were similar to those obtained with fatty acid in respect of systemic hyperlipæmia and deposition of absorbed fat. When fat was fed to human subjects along with lipase the systemic hyperlipæmia was much less than that normally found (Fig. 17, 21). On the other hand, if sodium cetyl sulphate, which inhibits the enzyme lipase, was fed to rats along with fat, fat absorption went on as usual and the picture was the same as for the absorption of neutral fat alone. Small quantities of fatty acid caused clumping of barium sulphate suspension in the intestinal lumen due to excessive mucus secretion, but glycerides did not. Delay in gastric emptying time was more marked after feeding fatty acids than glycerides. All these observations led Frazer to the limited, but important, conclusion that complete lipolysis was not an obligatory step in glyceride absorption. The older observation of Hunk and Rosenstein, that absorption of oleic acid may, under certain as yet undefined circumstances, give rise to particulate glyceride in the chyle, has been rediscovered by a number of investigators. Although this result differs from the observation of Frazer reported above, this does not in any way affect the general conclusion that pancreatic lipolysis need only be partial, since this fact has been fully confirmed by several independent workers using isotope labelled fats. Differences were also demonstrated between tributyrin and triolein. The former was more rapidly and completely hydrolysed *in vitro*, its contained fatty acids being water soluble; it was not emulsified in the intestinal lumen, it did not give rise to particulate fat in the cells, the chyle or the systemic blood, and it was not deposited in the fat depots. It was found to be absorbed normally by adrenalectomized animals. On the other hand, feeding triolein produced the changes already described as being characteristic of olive oil. Furthermore its absorption was depressed in adrenalectomized animals. Frazer concluded that partition between the oil and water phase was an important factor affecting the absorption and distribution of lipid molecules. The particulate oil phase passed preferentially in the chyle, while the

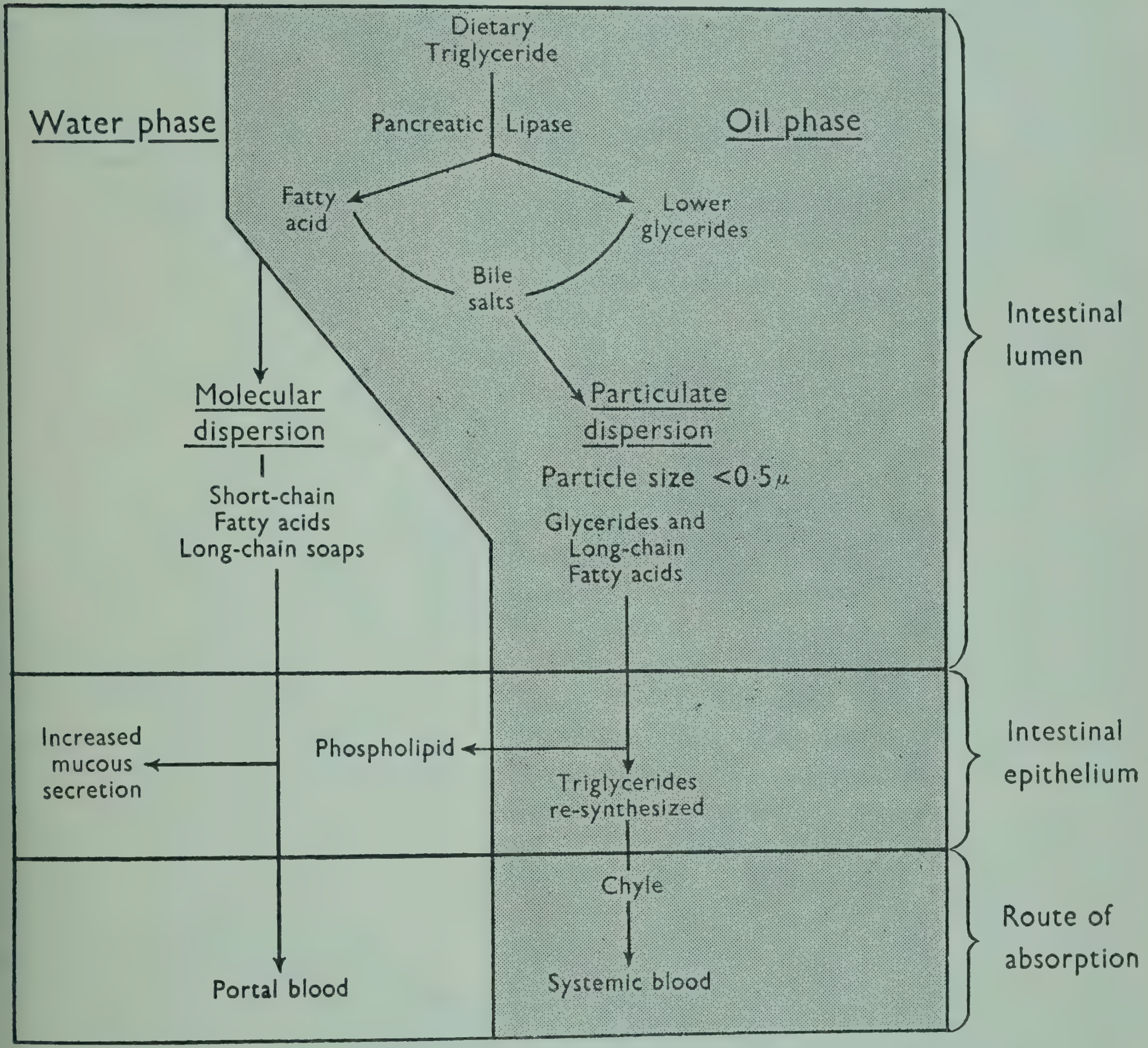

Dietary Triglyceride
Water phase
Pancreatic Lipase
Oil phase
Fatty acid
Lower glycerides
Bile salts
Molecular dispersion
Short-chain Fatty acids Long-chain soaps
Particulate dispersion
Particle size <0·5μ
Glycerides and Long-chain Fatty acids
Intestinal lumen
Increased mucous secretion
Phospholipid
Triglycerides re-synthesized
Intestinal epithelium
Chyle
Route of absorption
Portal blood
Systemic blood

FIG. 17, 19

The partition hypothesis of fat absorption (A. C. Frazer in Avery Jones, Ed. (1952), *Modern Trends in Gastro-enterology*. London: Butterworth).

water phase was distributed mainly by the portal route, due to the great difference in volume of portal blood and chyle flow. The differential distribution of long-chain and short-chain fatty acids has been amply confirmed by Chaikoff and his group, using isotope labelled materials.

Clearly the mechanism of absorption of insoluble glycerides is likely to differ from that of water soluble fatty acids. Frazer maintains that the partial hydrolysis of fats by pancreatic lipase gives rise to fatty acids and lower glycerides which, together with bile salts, form the system responsible for the production of a fine fat emulsion consisting of negatively charged particles less than 0·5 μ in diameter, with an interfacial film of fatty acid-bile salt-lower glyceride. Deficiency of pancreatic enzymes or bile from the intestinal lumen results in defective emulsification which can be corrected by the addition of the appropriate missing factor. The importance of emulsification in absorption is shown by the fact that unhydrolysable materials, such as paraffin, can be absorbed if they are finely emulsified before administration, or if they are fed as a mixture with olive oil, which provides the necessary emulsifying system for both. Paraffin alone is not hydrolysed, emulsified, or absorbed. The finely particulate fat may pass through fine canals in the outer border of the intestinal cell, which have been shown by Baker to be filled with lipid staining material during fat absorption (Fig. 17, 20 D and E).

The importance of intraluminar emulsification in glyceride absorption has been demonstrated in children with no pancreatic lipase in whom intraluminar emulsification is grossly defective. The administration of unemulsified fat to these patients causes no significant change in the quantity of particulate fat in the blood during the following six hours. If, however, finely emulsified fat is introduced into the duodenum, the particulate fat shows an increase, during the next two to three hours, that approximates to normal. The absorption of fat-soluble water-insoluble esters, with melting points above body temperature, is dependent upon the presence of an emulsifiable oil vehicle. It is possible that the main function of fine emulsification in the intestinal lumen is to bring the fatty material into intimate contact with the intestinal cell. Other factors, beside solubility in the oil phase and fine dispersion to a particle size of less than 0·5 μ and negative surface charge, may play a part in the passage of fat and fat-soluble molecules through the intestinal wall.

Comparison of the glycerides leaving the intestinal lumen and those found in the chyle show that triglycerides are resynthesized during passage through the intestinal wall. Paraffins also undergo considerable modification in the course of absorption. It is not known to what extent these molecular modifications are necessary steps in the absorptive mechanism. Chaikoff and his colleagues have shown from studies of ^{32}P turnover that phospholipid is not an obligatory intermediary in glyceride absorption. There is no doubt, however, that phospholipids are formed in the intestine and that they contain dietary fats. The exact reason for this is not known, but it may be that this phosphorylation provides an interfacial film which stabilizes fat droplets in the presence of the protein of the blood plasma. It was at one time thought that the adrenal cortex influenced fat absorption by controlling this phosphorylation. Recent work indicates, however, that the fat absorption defect in adrenalectomy is related to changes in water and electrolyte metabolism and particulate absorption.

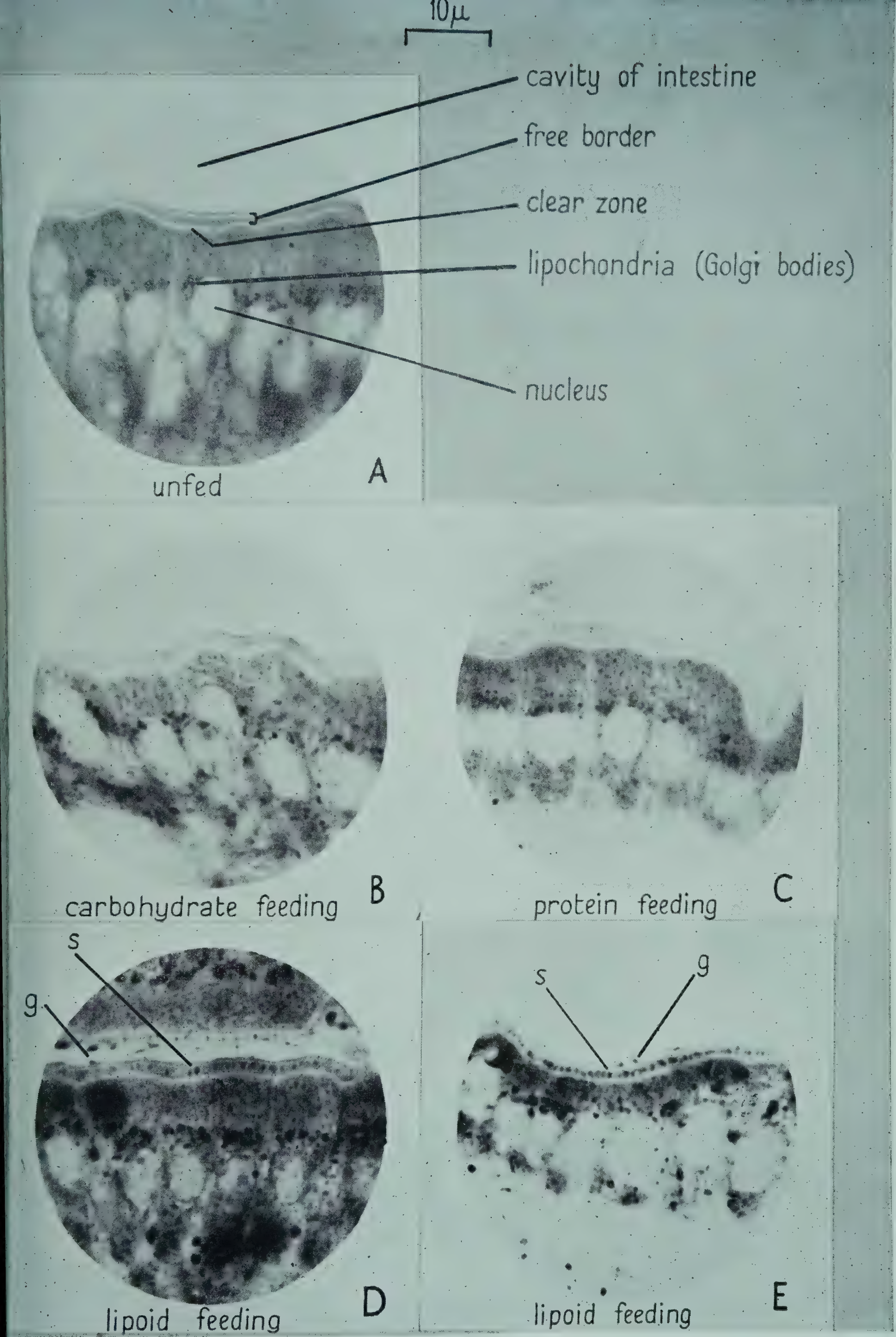

10μ
cavity of intestine
free border
clear zone
lipochondria (Golgi bodies)
nucleus
unfed
A
carbohydrate feeding
B
protein feeding
C
s
g
lipoid feeding
D
s
g
lipoid feeding
E

FIG. 17, 20 A–E

otomicrography of the intestinal epithelium of the unfed mouse, and of mice fed in various ays. (Gelatine sections coloured by the sudan black technique); *s*, lipoid spindle in free rder; *g*, lipoid body in cavity of intestine. (J. R. Baker (1951), *Quart. J. mic. Sci.* **92,** 86).

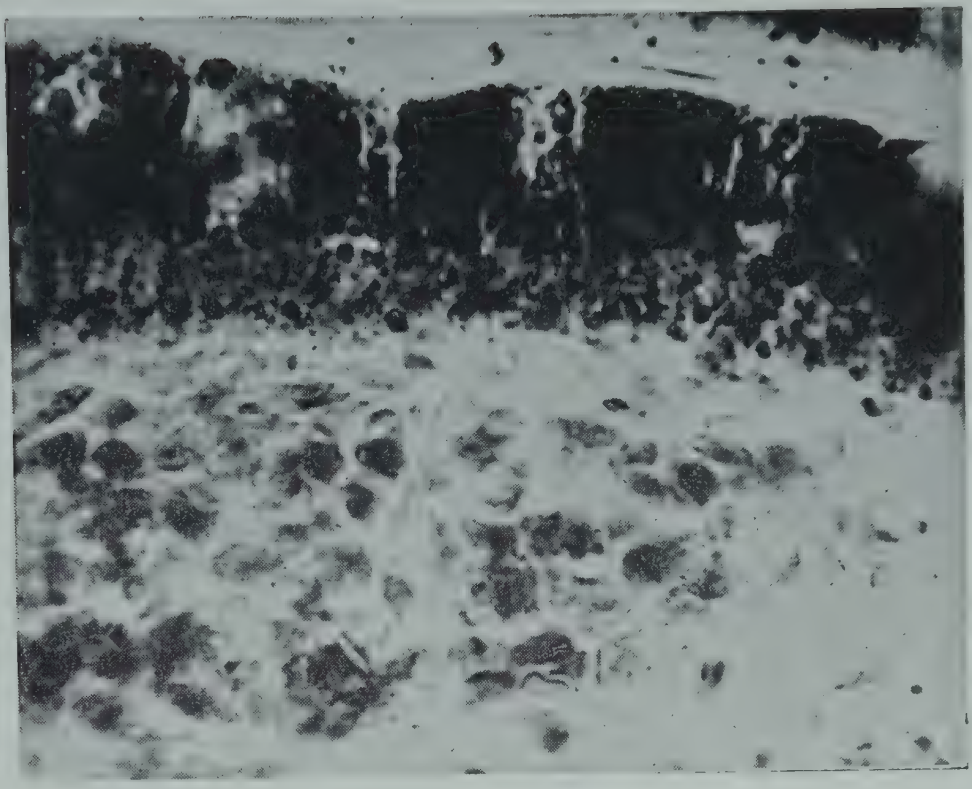


FIG. 17, 20F

Section of rat intestine stained Sudan IV and hæmatoxylin. The rat was given 1 ml. olive oil and 1 ml. water by intragastric tube. The animal was killed three hours later. The intestinal epithelium is loaded with fat: the centre of the villus is clear.

(By courtesy of A. C. Frazer.)

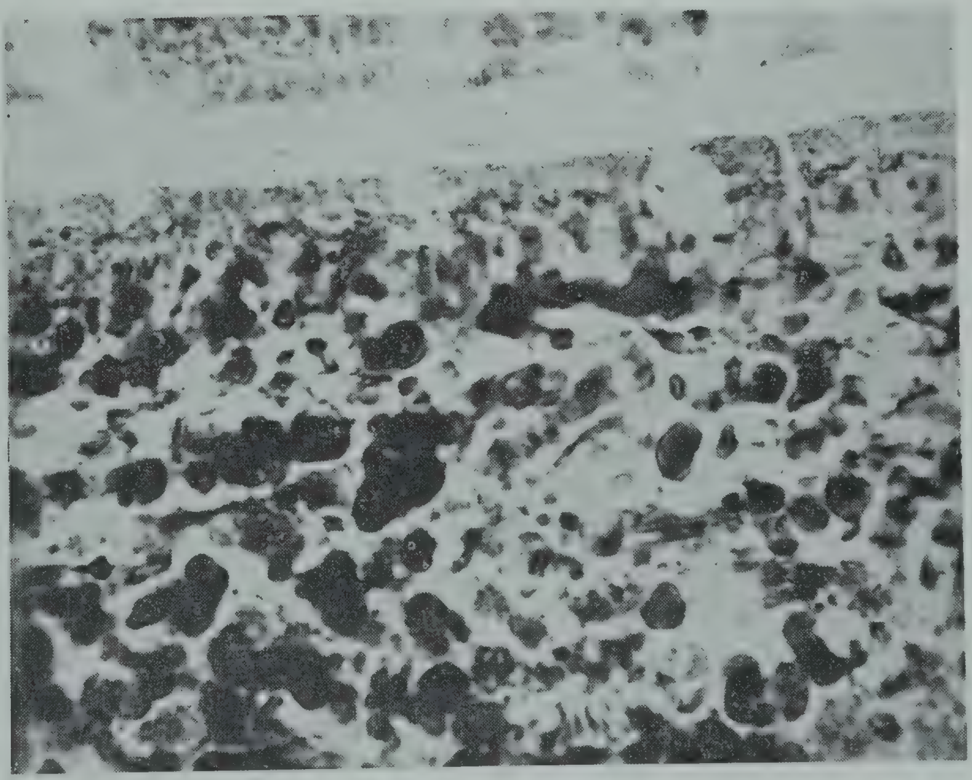


FIG. 17, 20G

Section of rat intestine stained Sudan IV and hæmatoxylin. 1 ml. olive oil plus 1 ml. water containing 5 mg. choline hydrochloride was given by intragastric tube. The animal was killed three hours later. The intestinal epithelium is largely clear of fat but the centre of the villus is now loaded with large fat droplets.

(By courtesy of A. C. Frazer.)

In rats, if choline, lecithin or glycerophosphate is administered at the same time as a fatty meal, the rate of fat absorption is increased by 50 per cent. or more and it appears to pass more quickly through the cells so that it may accumulate in the interior of the villus (Fig. 17, 20G). This is probably only one example of the way in which other dietary constituents may modify the absorptive functions of the intestinal cells. The proportion of fat that passes into the chyle, or into the portal blood, is therefore related to the type of fat being absorbed, the extent of hydrolysis, the amount of fatty acid that can be molecularly dispersed, and possibly the presence or absence of other dietary constituents.

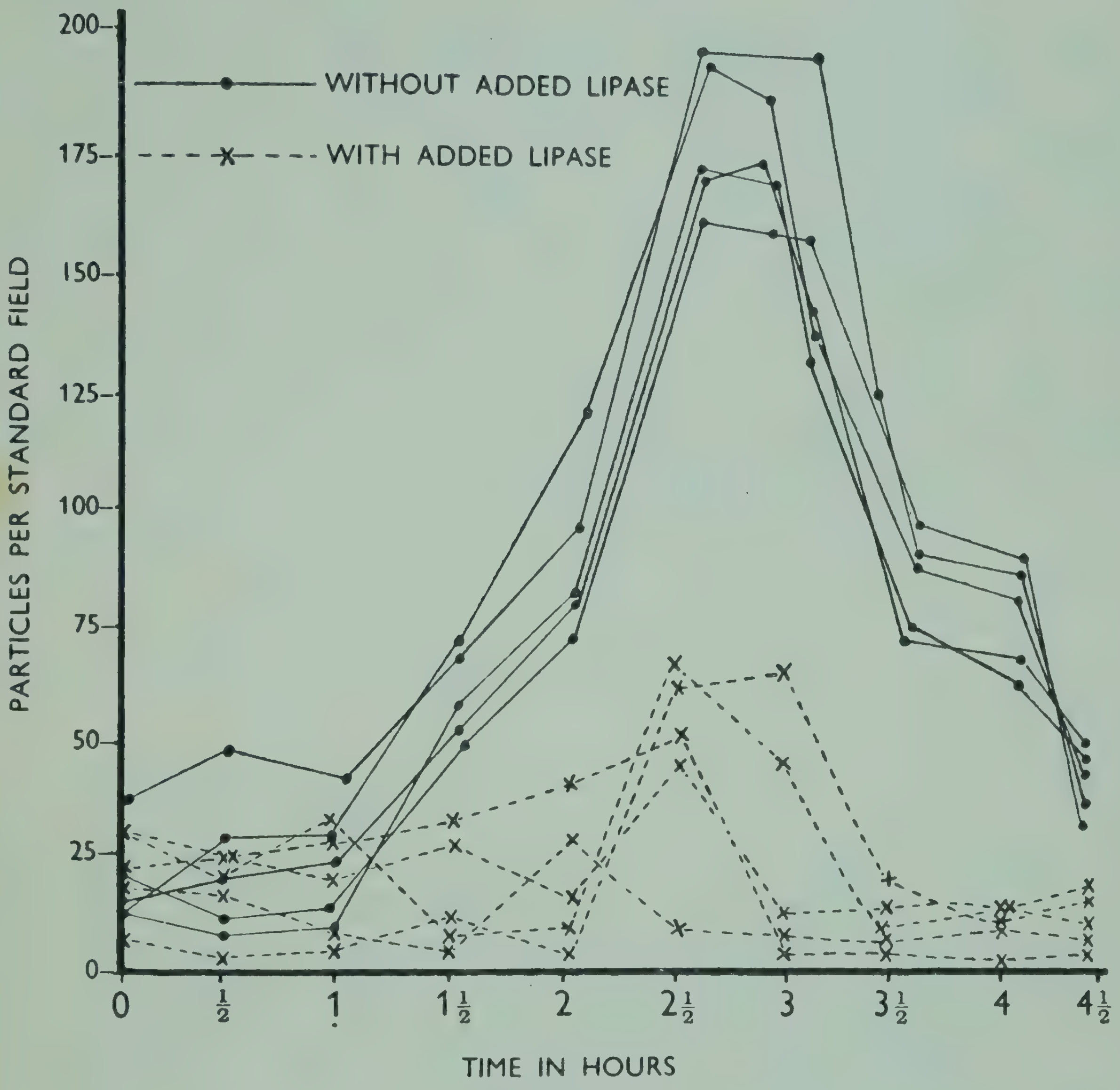


Fig. 17, 21

Five normal human chylomicrograms following ingestion of 30 g. neutral fat with and without added lipase (A. C. Frazer (1943), *J. Physiol.*, **102**, 329).

After passing through the intestinal cell long-chain triglycerides (with more than twelve carbon atoms) travel mainly through the lacteals to the thoracic duct, and so by way of the blood to the fat depots—shorter chain fatty acids, however, do not appear in the chyle and are thought to pass directly into the portal blood. Some reasons for this differential absorption have already been discussed. The outline of the process of fat absorption is indicated in Fig. 17, 19.

Fat absorption can be measured in man by estimating over a period of time the output of fat in the fæces when the patient is taking a diet of which the content of fat is known. Normally more than 95 per cent. of dietary fat is absorbed. The normal level of fat in the fæces is about 5 g. per day;

most of this is of non-dietary origin. When fæcal fat is increased it may be due to an increment in the dietary or the non-dietary element, or both. If fat absorption is to be accurately assessed, differentiation between dietary and non-dietary fat must be made. This can be done by altering the dietary fat load, or, better, by the use of labelled fat. The non-dietary fæcal fat may be derived from excreted fat, cellular debris or fat formed by intestinal organisms.

The intraluminar phase can be directly studied by feeding fat after passing a Miller-Abbott intestinal tube; the changes which fat undergoes in the lumen can then be examined in serial samples of intestinal contents.

The occurrence and extent of particulate absorption may be studied by the technique of counting chylomicrons. The subject consumes a standard meal containing a known amount of fat, and specimens of capillary blood are removed before the meal and at half-hourly intervals after it. The blood is allowed to clot and a drop of the serum is examined microscopically by dark-ground illumination. The fat droplets (chylomicrons) appear as bright particles showing Brownian movement and the number of these particles in each field of a standard size can readily be counted. For fasting blood this number is about 30, whereas at the peak of the lipæmia following a fat-containing meal the number may rise to 150 per field (Fig. 17, 21). Fat absorption is normally complete within 4½ hours after the ingestion of a normal quantity of fat, such as 30 g., and the maximum content of fat in the blood (*systemic lipæmia*), as demonstrated by the concentration of fat droplets or *chylomicrons*, usually occurs within two to three hours (Fig. 17, 21).

It has been shown by Chaikoff in America and by Borgström in Sweden, both using isotopically labelled palmitic and stearic acids, that long chain fatty acids appear in the chyle as triglyceride whether fed as acid or as triglyceride. On the other hand short chain fatty acids with fewer than ten carbon atoms appear more abundantly in the portal blood than in the chyle. These findings would suggest that the partition is determined rather by chain length than by extent of hydrolysis. Frazer's theory may therefore have to be modified accordingly.

Cœliac disease in children is a condition characterized by interference with weight gain and growth, and the passage of pale, bulky, offensive fluid motions that are usually fatty. It develops at weaning and is commonly associated with loss of appetite and abdominal distension. The cœliac child is irritable and unhappy. In severe cases, emaciation may be extreme and the child may die from intercurrent infection. Biochemical studies reveal depression of the absorptive functions of the small intestine, which is accompanied by radiographic changes indicating dilatation of the upper intestine, with loss of the normal mucosal pattern. Due to delayed absorption, fermentative changes may occur in the contents of the lower parts of the intestinal tract.

As a result of the clinical observation of Dicke, of Amsterdam, that his patients with cœliac disease showed steady improvement during periods of acute bread shortage at the end of the War, the importance of wheat flour in the ætiology of this condition has now been demonstrated. The main components of wheat flour are 85 to 90 per cent. of wheat starch and 10 to 15 per cent. of wheat gluten. If cœliac children are placed on a diet free from wheat gluten, they speedily return to normal, both clinically and biochemically. The re-introduction of wheat gluten into their diet causes a return of all the features of cœliac disease. Changes in dietary wheat starch, however, have no significant effect. Completely hydrolysed wheat gluten is not deleterious to cœliac patients; wheat gluten is well tolerated by normal children. It appears, on present evidence, that wheat gluten contains a peptide that cannot be satisfactorily handled by the cœliac child. This leads to depression of absorption from the small intestine and other effects, some of which are probably important in the establishment of the syndrome, since defective absorption alone does not account for many of its features.

The Absorption of Sterols and other Substances.—The intestine shows preferential absorption in dealing with the unsaponifiable fraction of the fats. Thus, cholesterol is absorbed readily, whereas related sterols derived from plants are not absorbed at all. Vitamins A and D are absorbed along with the fats. Carotenoids are absorbed by man without preference for any particular member of the group, whereas the cow absorbs carotene but not xanthophyll.

Vitamin K is absorbed in the presence of bile salts. In obstructive jaundice, when bile is excluded from the intestine, this vitamin is not absorbed and the plasma prothrombin and factor VII are lowered. There is then an increased clotting time and a tendency for hæmorrhage to occur (p. 149).

REFERENCES

Brocklehurst, R. J. (1950). Hormones of the digestive tract. *Advanc. Sci. Lond.* **26,** 212–223.

Code, C. F., Hightower, N. C. & Morlock, C. G. (1952). Motility of the alimentary canal in man. Review of recent studies. *Amer. J. Med.* **13,** 328–351.

Conference on triglyceride fats in human nutrition (1949). *Brit. J. Nutr.* **3,** 347–374.

Deuel, H. J., Jr. (1955). *The Lipids,* Vol. II. New York: Interscience Publishers.

Douglas, D. M. (1948). The activity of the duodenum. *J. Physiol.* **107,** 472–478.

Frazer, A. C. (1952). Fat metabolism. *Annu. Rev. Biochem.* **21,** 245–272.

Garry,, R. C. (1952). The nervous control of the movements of the alimentary canal. *Advanc. Sci. Lond.* **9,** 197–206.

Grossman, M. I. (1950). Gastrointestinal hormones. *Physiol. Rev.* **30,** 33–90.

Ivy, A. C. (1934). Physiology of the gall-bladder. *Physiol. Rev.* **14,** 102–139.

Ivy, A. C. (1947). Motor dysfunction of the biliary tract. *Amer. J. Rœntgenol.* **57,** 1–11.

Jones, Avery (1952). *Modern Trends in Gastro-enterology.* London: Butterworth.

Lipid metabolism (1952). *Biochem. Soc. Symp.* No. 8. Cambridge University Press.

Solomon, A. K. (1952). Electrolyte secretion in the pancreas. *Fed. Proc.* **11,** 722–731.

Thomas J. E. (1951). *The External Secretions of the Pancreas.* Oxford: Blackwell.

Verzar, F. & McDougall, E. J. (1936). *Absorption from the Intestine.* London: Longmans, Green & Co.

Youmans, W. B. (1949). *Nervous and Neuro-humoral Regulation of Intestinal Activity.* London: Interscience.

CHAPTER 18

THE LARGE INTESTINE

The Structure of the Large Intestine

THE general description of the intestine already given (Fig. 16, 1, p. 230) is applicable with certain modifications to the large intestine. The calibre of the large intestine is greater than that of the small, the transverse measurement being greatest at the cæcum and becoming narrower towards the rectum. The longitudinal muscle is gathered into three longitudinal strands which, since they are shorter than the other coats, produce a sacculated appearance (Fig. 18, 4). As it possesses no villi the mucous coat is smooth. The glands of the large intestine are long and very closely packed and are composed to a large extent of mucus-secreting goblet cells (Fig. 18, 3). Lymph nodules are present in the proximal part of the colon and especially in the vermiform appendix. Near the anus the mucous membrane changes to stratified squamous epithelium.

The size of the large intestine relative to the rest of the gut varies considerably in different species and is apparently related to the habitual diet of the animal. In the rabbit the cæcum is very large, with an appendix about the same size as a human vermiform appendix. In the horse the colon and cæcum are literally enormous and are concerned with the digestion of cellulose (p. 266). In carnivorous animals, on the other hand, the large intestine is relatively small. The cæcum and colon of man are intermediate in size; this might be expected, since his normal diet contains both flesh and vegetables. The capacity of the large bowel in man, as measured by a contrast enema radiologically controlled, is in the range 0·9 to 1·8 litres.

MOVEMENTS OF THE LARGE INTESTINE

The cæcum and colon usually fill with chyme from the small intestine when food is eaten, or when it reaches the duodenum, as the result of a peristaltic rush in the small intestine accompanied by a relaxation of the ileocolic sphincter. About four hours after it is taken a barium meal enters the cæcum and soon after this the meal quickly fills the ascending and transverse colon. Apparently this filling is almost completely passive, very little peristaltic movement being observed. On the average a barium meal reaches the hepatic flexure of the colon in six hours, the splenic flexure in nine hours and the pelvic colon in twelve hours, but there are considerable individual variations. Under ordinary conditions the time of passage of a meal through the intestine bears some relation to the composition of the meal, dry meals tending to pass more slowly. All observations indicate that foods having larger residues pass more rapidly, a fact that is important in dietetics.

Nerve Supply of Colon and Rectum.—It can be shown that fibres from the vagus reach the proximal colon. Section of the vagi leads to degeneration of nerve fibres in the proximal and intermediate colon but not in the distal colon. The outflow from the spinal cord to the distal colon in man is described in Fig. 18, 1. The thoraco-lumbar sympathetic outflow leaves the lumbar sympathetic chains, forming plexuses on the aorta, and passes eventually in the mesentery to the colon as the inferior mesenteric nerves. The 'presacral nerve,' usually numerous nerve filaments, emerges from the plexuses in front of the aorta and joins the hypogastric plexuses at the side of the rectum and bladder. The lowest part of the hypogastric plexus is called the pelvic plexus. The main sacral (parasympathetic) outflow in man (pelvic splanchnic nerves) comes from the second, third and fourth sacral nerves and joins the hypogastric plexus. The parasympathetic outflow joins the hypogastric plexus before arriving at the rectum and internal sphincter. The striated muscle of the external sphincter is supplied by the pudendal nerve (Figs. 18, 1, and 18, 2).

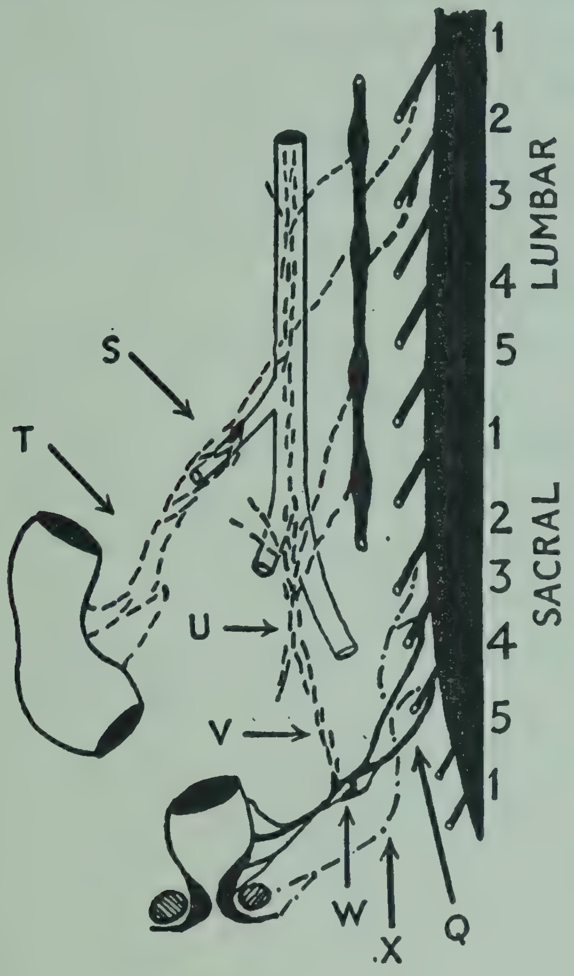


FIG. 18, 1

The nervous outflow to the distal colon in man (R. C. Garry (1934), *Physiol. Rev.* **14,** 107).

Q = pelvic splanchnic nerves.
S = inferior mesenteric plexus.
T = inferior mesenteric nerves.
U = presacral nerve.
V = hypogastric nerve.
W = hypogastric plexus.
X = pudendal nerve.

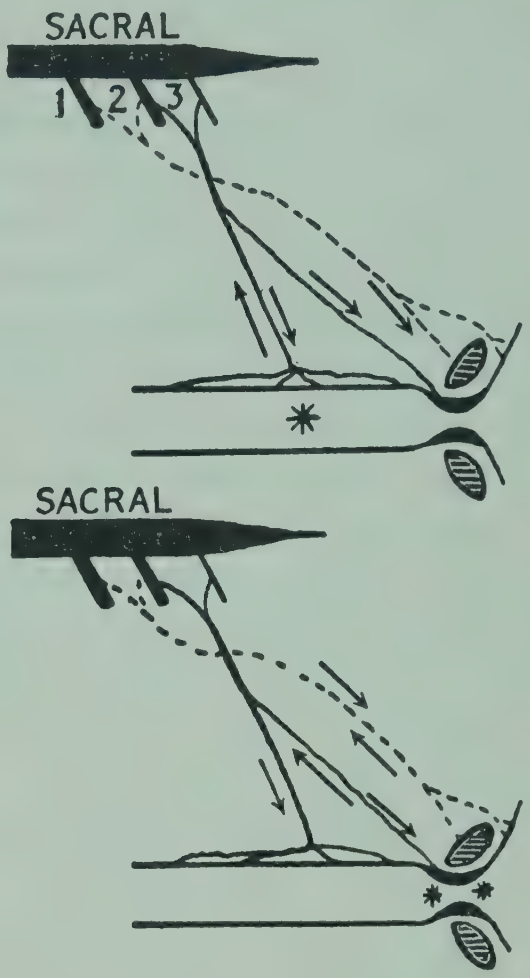


FIG. 18, 2

Diagram of the supposed course of the nerve impulses in defæcation. The sacral cord of the cat: continuous line the pelvic nerve, interrupted line the pudendal nerve; * the point of stimulation. (*After* R. C. Garry (1933), *Trans. roy. med.-chir. Soc. Glasgow*, in *Glasgow med. J.*, **28,** 9, Fig. 6, Nov. and Dec. 1933.)

The lumbar outflow (sympathetic) is wholly inhibitory to the colon.

When it is cut, or when the cord is anæsthetized, the colon is more active and the internal anal sphincter relaxes. The lumbar outflow has been divided in man in an attempt to relieve the symptoms of megacolon and chronic constipation. After experimental section of the sacral (parasympathetic) outflow in animals the movements of the distal colon become weak, with contraction of the internal anal sphincter. In animals megacolon may develop.

The effect of stimulation of the sympathetic nerves depends on the experimental conditions and on the state of the colon—if the tone of the colon is high inhibition is the usual result, but if the tone is low sympathetic stimulation may increase its activity. Stimulation of the parasympathetic increases the muscular activity of the colon. In dogs there are periodic alterations in the tone of the proximal parts of the colon, though the distal segment does not behave in this way.

Spinal Cord and Defæcation.—Throughout the greater part of its length the intestinal tract shows a large degree of autonomy. Coordinated movements continue after complete destruction of the spinal cord. At its caudal extremity, however, the gut comes directly under control of the nervous system. If the lumbo-sacral region of the cord is destroyed the anal sphincters, the internal sphincter of smooth muscle and the external sphincter of striped muscle, are patulous and the caudal portion of the large intestine is paralysed. In time there may be some return of activity and the internal sphincter may recover tone. The caudal region of the gut may thus eventually recover its autonomy. Under those conditions fæcal matter may be expelled at intervals but the patient is unaware of it and can exert no voluntary control.

If the spinal cord be transected craniad to the sacral region of the cord the condition of the gut, so long as spinal shock persists, resembles that seen after destruction of the sacral region of the cord. Subsequently, when the spinal shock has passed off, reflex defæcation is established. Expulsion of fæcal matter may be complete but the patient is unaware of the act and is unable to exert any voluntary control over it.

The reflex pathways subserving defæcation are shown in Fig. 18, 2. Stimulation of the distal colon in the cat causes contraction of the colon through the pelvic splanchnic nerve and relaxation, through the pelvic splanchnic and pudendal nerves, of the internal and external sphincters respectively. When the fæces pass through the sphincter these contractions and relaxations are reinforced by reflexes involving afferents and efferents in both pelvic splanchnic and pudendal nerves.

Defæcation is a complex reflex act involving not only contraction of the rectum and caudal portion of the colon, inhibition of both internal and external sphincters but also contraction of the diaphragm and of the skeletal muscles of the abdominal wall. The levator ani muscles in the pelvic floor also aid defæcation by pulling up the anal canal over the fæcal mass. During defæcation the pressure in the rectum may rise as high as 200 mm Hg during violent expulsive efforts. Such high intra-abdominal pressures must have important effects on the circulation, especially on the venous return to the heart.

Three or four times a day a massive peristaltic wave pushes the contents of the ascending colon onwards so that the fæces accumulate proximal to the pelvirectal flexure, leaving the rectum empty. One of these massive movements may push the contents into the rectum and the subject at once experiences a sensation of fullness in the rectum and a desire to defæcate. Such a peristaltic movement frequently follows the eating of a meal and is thought to be due to filling of the stomach (gastrocolic reflex of Macewen). At the same time the mucosa of the colon becomes red and the secretion of mucus is increased. This reflex is the basis of the usual habit of defæcating after a meal, very often after breakfast. There is no doubt, however, that the time of defæcation depends very much on habit and the opportunities available according to the occupation of the individual. Most adults defæcate once per day but three times a day may be perfectly normal, and so may once in three days. If the human being refrains from the act of defæcation, either by voluntary contraction of the external sphincter of the anus or, possibly, by voluntary inhibition of the rectum through the pelvic nerves, then the sensation rapidly fades. In this fashion the habit of constipation may only too easily be acquired.

Distension of the lower colon in man by means of a balloon produces a sensation of colic in the abdomen. The nerve impulses pass into the spinal cord along the sympathetic pathways. Sudden distension of the rectum on the other hand produces a desire to pass wind or motion and a reflex contraction of the external sphincter, the afferent pathway being in the pelvic nerves (Fig. 18, 2). The rectum is able to distinguish between flatus and fæces probably because flatus produces less distension. When these sensations rise to consciousness a voluntary decision is made as to whether the contraction of the external sphincter, initially reflex, should be maintained or relaxed.

ABSORPTION AND SECRETION IN THE LARGE INTESTINE

About 350 g. of chyme passes from the small intestine through the ileocolic valve into the cæcum each day. This fluid material is greatly reduced in bulk during the thirty-six hours of its stay in the colon by absorption of the greater part of the water mainly in the cæcum and ascending colon. The retention of this water in the body is of importance for the maintenance of its fluid equilibrium. If diarrhœa, that is the passage of frequent watery stools, persists for more than twenty-four hours dehydration (Chapter 35) occurs which may, especially in infants and young children, be quickly fatal. Although it does not do so to any great extent under normal conditions the mucosa of pelvic colon can absorb water and solutes of low molecular weight. Advantage is frequently taken of this fact when fluid must be introduced into the body but the patient cannot take water by the mouth, for example after surgical operations on the upper alimentary tract. Some five or six litres per day of fluid may be administered per rectum for two or three days, but it is important to remember that the absorptive properties of the distal colon are strictly limited and that, while many drugs can be introduced into the body by this route, strongly hypertonic and colloidal (protein) solutions are not absorbed. Fat is not absorbed. Chloride ions are more rapidly absorbed than sodium ions. Since absorption in the proximal colon is better than in the distal colon it may be advantageous to introduce nutrient enemata through a long tube reaching well up in the large bowel. It is possible that most of the absorptive powers attributed to the colon are due to the passage of the substance to be absorbed through the ileocolic valve into the lower ileum.

An isolated loop of colon secretes a small amount of mucus which is produced by the goblet cells; mechanical stimulation arouses a small quantity of watery secretion with practically no enzymatic activity. The mucus lubricates the fæces and neutralizes any acids formed in them. The pH of the contents of the colon is about 8.

If one pelvic splanchnic nerve is cut, stimulation of either the peripheral or the central end produces secretion of mucus in the large intestine. The former is a direct effect, the second a reflex effect. Secretion can also be obtained by injection of vagomimetic drugs such as pilocarpine and can be prevented by atropine. It used to be thought that the colon excreted metallic salts, but due allowance was not always made in the older experiments for incomplete absorption in the small intestine of salts given by the mouth. If bismuth is given subcutaneously it is excreted by the large intestine and stains the mucosa black. However, if iron or calcium are given intravenously the output of these by the large bowel is not increased. Large amounts of potassium and other electrolytes can be removed from the body by the introduction of ion exchange resins (p. 38) into the lower bowel.

The Bacteria of the Intestine

The intestine of the new-born child is bacteriologically sterile and the material excreted from it is a semifluid greenish mass known as *meconium*. A few days after birth, however, the intestine is invaded by bacteria which remain throughout life. Normally the stomach and small intestine contain only a sparse transient bacterial flora, the organisms being destroyed by the gastric hydrochloric acid and by the antibacterial mechanism of the small intestine. Even when the stomach contains large numbers of organisms the small intestine contains very few. Enormous numbers of organisms of all kinds are found in the colon of which they must be regarded as normal inhabitants. They bring about fermentation of carbohydrates with the production of gases such as hydrogen, carbon dioxide and methane. A large part of the gas found in the colon is, however, derived from the atmospheric air which passes surprisingly quickly down the alimentary canal. Air which was swallowed has been seen by X-rays to reach the cæcum in 14 minutes and to be passed as flatus in 30 minutes. Part of these gases is absorbed and the remainder, about 500 ml. per day, is passed per rectum as flatus. The bacterial flora may digest a small amount of the cellulose found in the rougher parts of fruits and vegetables but this process is of very minor importance in man though of great importance in certain herbivorous animals. Cellulose is not attacked significantly by any of the enzymes of the human intestine. The colonic bacteria also act upon proteins and amino-acids, with the production of amines and phenols by reactions involving mainly decarboxylation and deamination (p. 350). Among the materials produced are indole and skatole (p. 349). Probably the most important function of the intestinal flora in man is to produce vitamins, particularly those of the B complex. It is likely that bacterial

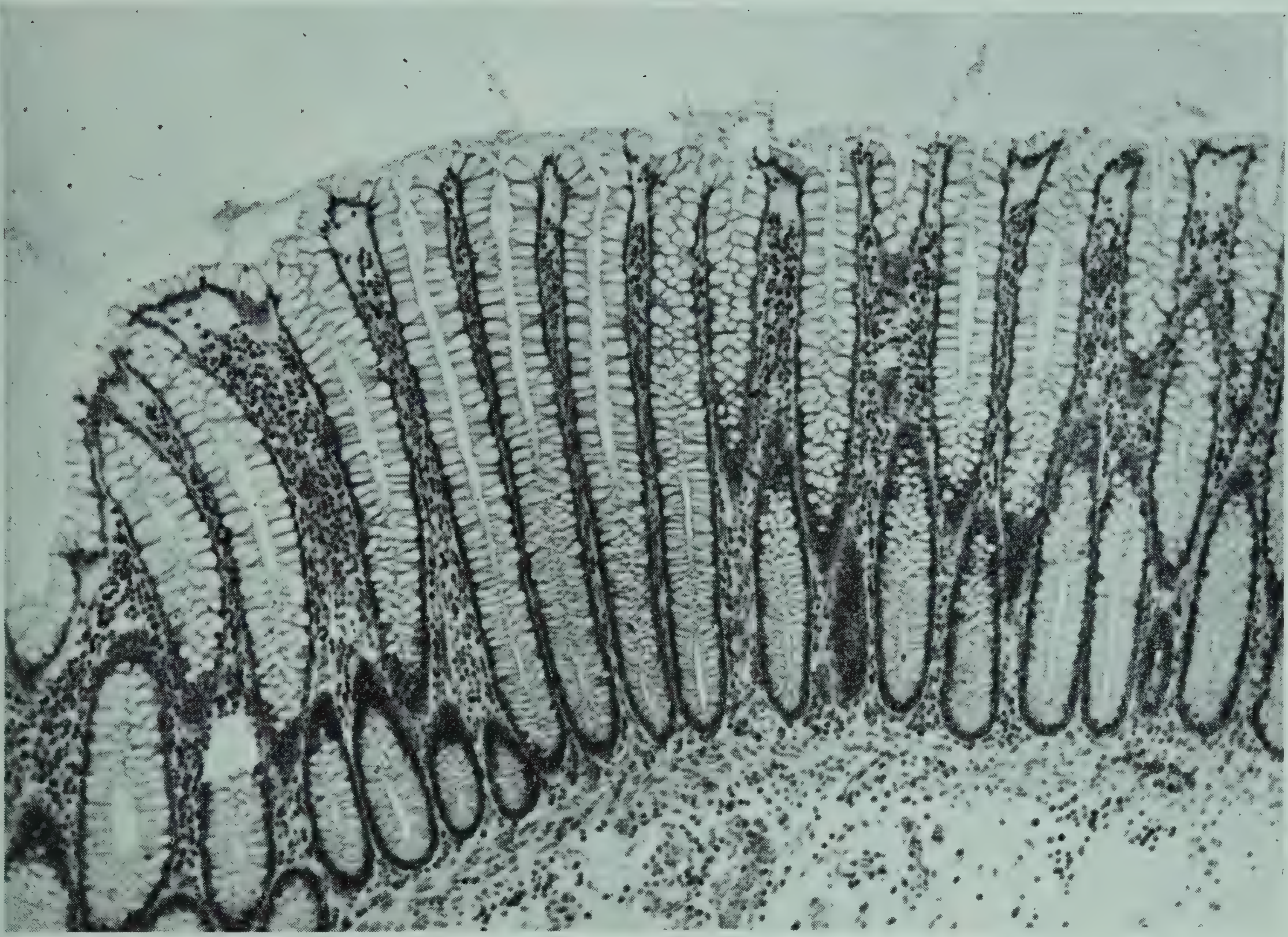


FIG. 18, 3

Transverse section of human large intestine showing large numbers of mucus-secreting cells. Hæmalum and eosin; × 115.

(*By courtesy of A. C. Lendrum.*)

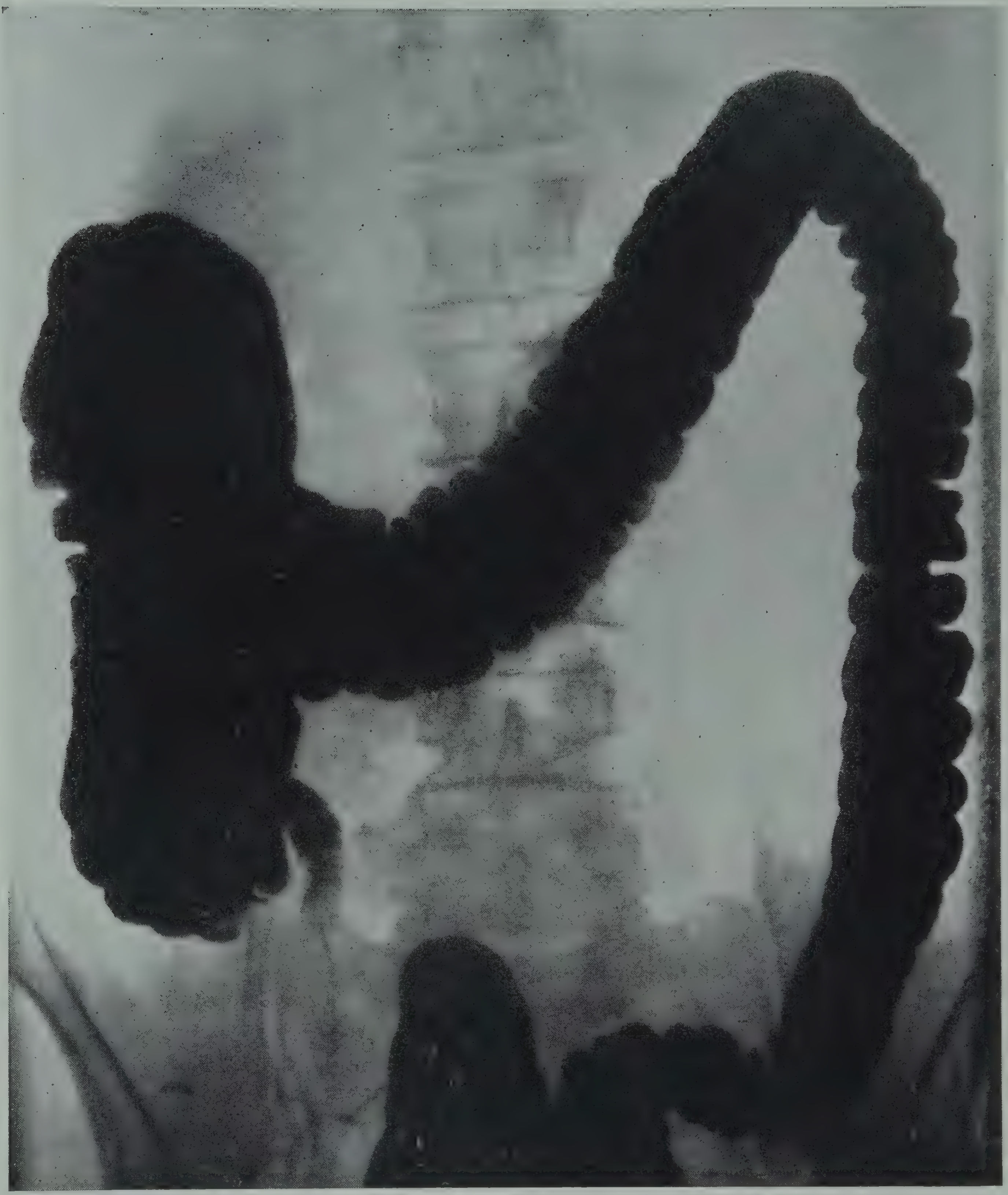


FIG. 18, 4

X-ray photograph of the human colon outlined by barium sulphate introduced in the form of an enema. A small amount of barium has entered the ileum. The indentations of the column of barium, visible particularly in the descending colon, are known as *haustrations*. They are the result of muscular contraction and vary greatly both in number and form from time to time.

synthesis of vitamins in the lower alimentary tract is responsible for fulfilling a substantial, and certainly a nutritionally important, proportion of the daily requirements of riboflavin, nicotinic acid, biotin, folic acid, and vitamin K. Relatively large amounts of vitamin B_{12} are also synthesized in the large intestine but are not absorbed.

THE FÆCES

At the end of about thirty-six hours the contents of the large intestine have become solid or semisolid and are known as *fæces*. The amount produced in twenty-four hours varies widely, according to the diet, from about 80 to 200 g., but is usually about 100 g. The reaction is usually slightly alkaline on the surface (pH 7·0 to 7·5). The *colour* is due to the presence of the pigment *stercobilin* derived from bile pigment (p. 415), but in disease the colour may be greatly changed, for example to black in the presence of altered blood or to clay-coloured in the absence of bile. The fæcal *odour* is due mainly to the presence of indole and skatole (p. 349) and to gases produced by fermentation of carbohydrate.

The water content of the fæces varies from 60 to 80 per cent. according to the length of time they are retained in the colon. In health, and on a relatively normal diet, the proportion of the daily fæcal mass attributable to undigested or unabsorbed food is small. The fæces consist chiefly of the residues of bile and other intestinal secretions, mucus, leucocytes and desquamated epithelium and enormous numbers of bacteria. The composition of fæces is as follows:

(1) Inorganic material. This accounts for 10 to 20 per cent. of the total solid and consists mainly of calcium and phosphates.

(2) Nitrogenous material equivalent to 1 to 2 g. of nitrogen. A large proportion of this, comprising as much as one-third of the total solids, is bacteria. The remainder consists of intestinal secretions, undigested proteins (normally present in very small amounts) and cellular debris.

(3) Fatty material extractable by ether. The fat content of fæces varies greatly with the amount of fat in the diet; with a small fat intake, in the range 50 to 70 g. for example, the absorption of fat is practically complete (95 per cent.). With the average fat intake approximately 10 to 20 per cent. of the total solids consists of fatty acids, neutral fats, phospholipids and sterols. Normally about 75 per cent. of the fatty material is in the form of so-called 'split fat,' that is free fatty acids. The remaining 25 per cent. includes neutral fat derived from cellular debris, from bacteria, and to a certain extent from the food. It also contains sterols such as coprosterol, produced by the reduction of cholesterol, and the plant sterols, which are normally not absorbed. In cœliac disease and sprue, fat may form as much as 70 to 80 per cent. of the dried fæcal matter and in these conditions the fæcal mass is much increased in volume.

(4) Undigested material. A variable proportion of the fæcal solids consists of cellulose and other undigested food residues such as the seeds and skins of fruits.

The composition of the fæces depends very little on the constitution of the diet, except that when the food contains much insoluble indigestible material, such as cellulose, the bulk of the intestinal contents is greater; intestinal peristalsis is stimulated so that the fæces pass more quickly through the large intestine and less time is allowed for concentration. It may be that one of the important advantages of a cellulose-containing vegetable diet is this stimulating action and people with infrequent bowel action may benefit by taking a diet which contains a certain amount of roughage. The effect of increasing roughage can be seen in some figures quoted from Rubner.

TABLE 18, 5

Kind of Bread	Wt. of moist fæces, g.	Wt. of dried fæces, g.	Percentage of Food	Nitrogen, g.
From fine flour	133	25	4	2·2
From coarse flour	253	41	7	3·2
Brown bread	318	76	12	3·8

Fæces continue to be formed during starvation; though somewhat smaller in quantity their composition remains essentially unaltered. This shows clearly that the fæces are produced chiefly within the intestinal tract and are not simply unabsorbed residues of food-stuffs.

REFERENCES

CREGAN, J., DUNLOP, E. E. & HAYWARD, N. J. (1953). The bacterial content of human small intestine in disease of the stomach. *Brit. med. J.* ii, 1248–1251.

GARRY, R. C. (1934). The movements of the large intestine. *Physiol. Rev.* **14,** 103–132.

GOLIGHER, J. C. & HUGHES, E. S. R. (1951). The sensibility of the rectum and colon. *Lancet,* i, 543–548.

GRACE, W. J., WOLF, S. & WOLFF, H. G. (1951). *The Human Colon.* London: Heinemann.

CHAPTER 19

INTERMEDIARY METABOLISM METHODS OF STUDY

THE study of intermediary metabolism aims at an understanding of the changes undergone by substances in the process of their utilization by the body and the reactions in which they take part after their absorption. The problems involved have been attacked in several ways.

1. If we know the chemical constitution of the substance under examination we can work out theoretically the possible changes which its molecule might undergo in the body. We can then feed to an animal or to man all the theoretically possible intermediate substances from the list to see whether they are made use of as readily as the test substance. If they are, then they *may* be intermediary metabolites of the substance under examination. For example butyric acid is oxidized more rapidly than β-hydroxybutyric acid and presumably therefore butyric acid is not oxidized by way of the hydroxy derivative.

2. We can examine the urine for some substance which increases in amount when the material under investigation is administered and which decreases when administration ceases. For example the amount of urea in the urine increases after the administration of amino-acids (p. 335). It is concluded therefore that at least part of the urea nitrogen is derived from the nitrogen of amino-acids.

3. We can feed excessive amounts of the test material to an experimental animal and examine the urine for abnormal constituents which might be intermediary breakdown products. For example, administration of excessive amounts of fats is followed by excretion of acetoacetic acid in the urine (p. 332). This method is open to the objections that the substance found in the urine may be a by-product not on the direct line of metabolism and that the loading of the body with unusually large amounts of the test material may involve abnormal routes of metabolism.

4. We may perfuse an isolated organ, such as the liver, with the material under investigation and examine the perfusion fluid for breakdown products. For example, when leucine is examined in this way the acetoacetic acid content of the perfusion fluid rises. Hence the oxidative deamination of leucine proceeds by way of acetoacetic acid (p. 332), and leucine is a ketogenic amino-acid.

5. The test substance may be incubated with finely mashed tissue (*tissue brei*) or with thin slices of an organ such as liver or kidney, and the concentration of a possible breakdown product estimated. Thus the production of urea by slices of liver tissue is increased when ornithine is added (p. 343), so that the latter is presumably an intermediate compound in the formation of urea from amino-acids.

6. We may introduce a 'signpost' or marker group into the molecule to aid in the subsequent identification of breakdown products. Knoop, for instance, introduced the phenyl group into fatty acid molecules to facilitate the study of the progress of their oxidation (p. 329). This

method suffers from the drawback that molecules so marked may be treated by the body as foreign substances and may not be dealt with in the same way as natural compounds.

7. The role played by a particular substance may in some cases be ascertained by feeding to an experimental animal a diet deficient in that substance. Thus when diets deficient in nicotinic acid (p. 127) are fed to animals it is found that the amount of nicotinamide nucleotides (p. 61) in the blood and tissues falls below the normal level.

8. Measures involving surgical operations on experimental animals have been employed to study intermediary metabolism. London and his pupils in Leningrad devised the technique of angiostomy, which involves the insertion of cannulæ leading to blood vessels so that blood may be withdrawn or material may be injected at any time. Dogs with cannulæ to the hepatic, portal and renal veins have lived apparently normal lives for many months. Again, in the Eck fistula operation a communication is made between the portal vein and the inferior vena cava. By this means the liver is largely short-circuited and the portal blood from the small intestine flows into the inferior cava. It is thus possible to investigate the part played by the liver in the intermediary metabolism of many substances.

A more drastic measure is removal of the whole liver (hepatectomy), employed by Mann and his colleagues. During the short time for which animals survive this operation urea formation ceases. The liver thus appears to be the site of urea formation in the intact animal (p. 342). Similar information may be obtained by the use of liver poisons, such as chloroform, carbon tetrachloride, or phosphorus, which damage the liver cells.

9. In a somewhat similar way drugs are sometimes administered to interfere with the normal metabolic processes. The glycoside phlorrhizin, by poisoning the renal tubules and preventing reabsorption of glucose, produces severe glycosuria (p. 305). It has proved to be of great service in investigating the intermediary metabolism particularly of carbohydrates, and in determining which of the many amino-acids can give rise to glucose after deamination.

10. Certain individuals are met with who fail to carry out normally some metabolic process. Such 'inborn errors of metabolism,' as they have been called by Garrod, are rare, usually both congenital and familial, and are as a rule quite compatible with health and a normal span of life. This group of metabolic disorders includes alkaptonuria, in which tyrosine metabolism is abnormal (p. 348), cystinuria, in which the metabolism of cystine is deranged (p. 355), and pentosuria (p. 305). From the detailed study of patients with such disorders much has been learnt about metabolic pathways.

11. It frequently happens that disease in man can be made to throw light on physiological processes. The study of diabetes mellitus (p. 307) and of glycogen disease (p. 299) has thus yielded much information about the normal processes of carbohydrate metabolism.

12. **Isotopes.**—Isotopes are elements with the same atomic number, that is the same nuclear charge, but with different atomic weights. Their use in the study of intermediary metabolism has recently provided the most satisfactory method for tracing the history of a given atom from its introduction into the body to its elimination.

Stable Isotopes.—As an example of naturally occurring stable isotopes we may take carbon. Stable carbon atoms of two types occur naturally (Fig. 19, 1). In one form the nucleus consists of 6 protons, each with one positive charge, and 6 neutrons which are electrically neutral. The nucleus therefore carries 6 positive charges, which are counterbalanced by the negative charges on 6 planetary electrons. The whole atom is thus electrically neutral. The atomic weight is 12 and is derived from the protons and neutrons of the nucleus, the mass of the planetary electrons being negligible. In the other type of carbon atom the nucleus consists of 6 protons and 7 neutrons. It likewise carries 6 positive charges, which are balanced by the negative charges of 6 planetary electrons. Because it contains 13 protons and neutrons its atomic weight is 13.

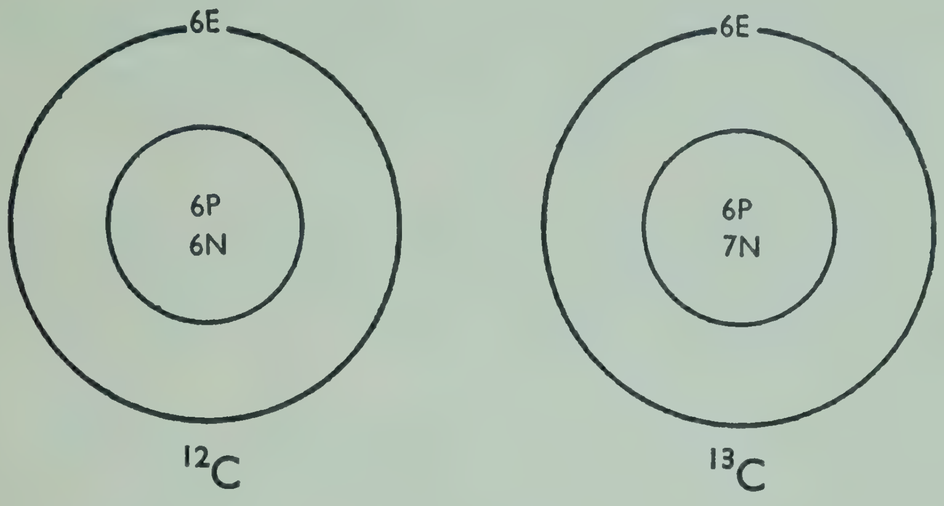


FIG. 19, 1

Stable Isotopes of Carbon.

^{12}C has a nucleus consisting of 6 protons (P) and 6 neutrons (N) and an outer shell of 6 electrons (E). ^{13}C has 6 protons and 7 neutrons in the nucleus and 6 planetary electrons.
Both have a nuclear charge, and therefore an atomic number, of 6, but their atomic weights are $6+6=12$ and $6+7=13$ respectively.

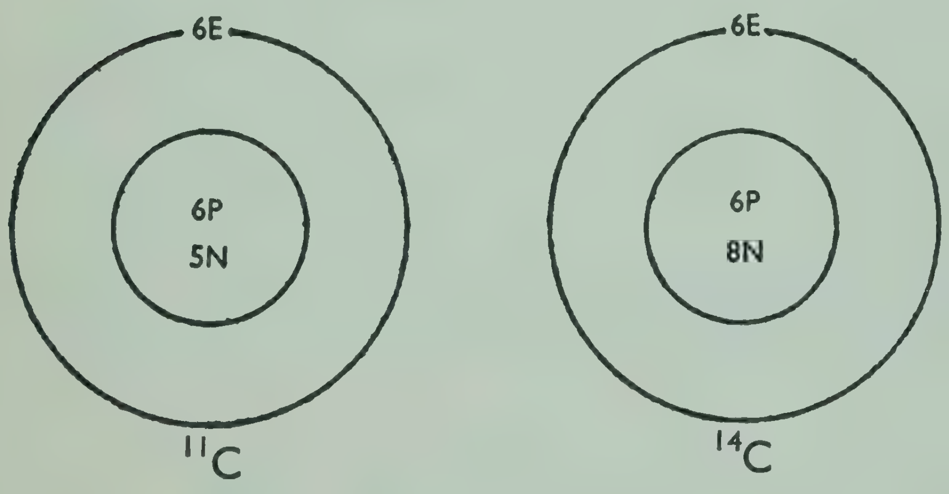


FIG. 19, 2

Radioactive Isotopes of Carbon.

^{11}C has a nucleus containing 6 protons and 5 neutrons, that is an atomic weight of 11.
^{14}C has a nucleus containing 6 protons and 8 neutrons, that is an atomic weight of 14.
Both ^{11}C and ^{14}C have a nuclear charge and therefore an atomic number of 6, and both have 6 planetary electrons.

Both types of carbon atom have a nuclear charge of 6 and therefore an atomic number of 6, but one has an atomic weight of 12 and the other an atomic weight of 13. They are represented in symbols thus:

$$^{12}C_6 \quad \text{and} \quad ^{13}C_6$$

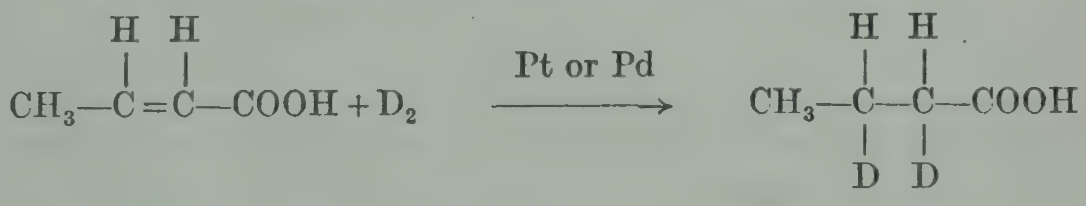
where the subscript represents the atomic number or nuclear charge, and the superscript represents the atomic weight. As all carbon atoms have an atomic number of 6 the subscript may be omitted and the two types of carbon atom represented simply by

$$^{12}C \quad \text{and} \quad ^{13}C$$

Carbon as it occurs in nature is a mixture of the two isotopes with ^{12}C preponderating. In fact 99 atoms of ^{12}C are found for every atom of ^{13}C. The mean atomic weight is therefore nearer 12 than 13 and is known to be 12·005.

Similarly chlorine, as ordinarily prepared, is a mixture of two isotopes, ^{35}Cl and ^{37}Cl, with the former preponderating, and the atomic weight is therefore between 35 and 37 but nearer 35. It is in fact 35·46.

Hydrogen exists also in the form of isotopes. The commoner type of hydrogen atom has a nucleus of one proton and has one planetary electron. Its atomic weight is 1 and its atomic number is 1. It is usually written as ^{1}H, or simply H. Hydrogen also exists in a 'heavy' form in which the nucleus containing one proton and one neutron carries one positive charge which is counterbalanced by the charge on the single planetary electron. This form is known as *heavy hydrogen* or *deuterium*. Its atomic weight is 2 and its atomic number is 1. It is usually represented as ^{2}H or ^{2}D, or simply as D. Ordinary water contains a very small proportion of its hydrogen (0·02 per cent.) in the form of deuterium, but this deuterium-containing water, or 'heavy water,' can be concentrated by fractional distillation. When heavy water is electrolysed, deuterium gas is given off. Organic compounds containing D in place of H may then be synthesized. For example, butyric acid containing two atoms of deuterium can be prepared by treating crotonic acid with deuterium gas in the presence of a platinum, palladium, or nickel catalyst.

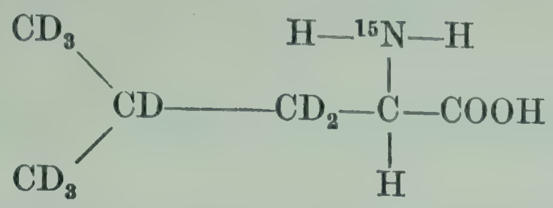


$$CH_3\text{—}\overset{H}{\overset{|}{C}}=\overset{H}{\overset{|}{C}}\text{—}COOH + D_2 \xrightarrow{\text{Pt or Pd}} CH_3\text{—}\underset{D}{\underset{|}{\overset{H}{\overset{|}{C}}}}\text{—}\underset{D}{\underset{|}{\overset{H}{\overset{|}{C}}}}\text{—}COOH$$

Such a compound may then be given to animals and its fate determined by following up the D atoms. Or heavy water (D_2O) suitably diluted with ordinary water (H_2O) may be injected into an animal and the chemical constituents of the tissues may later be isolated and separately analysed for D. The water formed on combustion is carefully collected and its D content determined by measurements of density or refractive index.

The nitrogen isotope ^{15}N, which is found to the extent of 0·368 per cent. in ordinary nitrogen, has, in the hands of Schoenheimer, Rittenberg, and others in America, been of immense value in the study of intermediary protein metabolism. Ammonium salts containing ^{15}N may be used in the synthesis of labelled amino-acids. The concentration of ^{15}N can be estimated subsequently in the nitrogen of the tissues or excreta by the *mass spectrometer*.

It is, of course, possible for a molecule to be labelled with more than one type of isotope. For example, leucine containing both D and ^{15}N has been used by Schoenheimer.



Such labelled compounds are treated by the animal body in exactly the same way as are their normal analogues, and their application in the study of intermediary metabolism will be illustrated later by specific examples.

Radioactive Isotopes.—Some isotopes, which do not occur naturally but may be prepared artificially, are radioactive, that is they disintegrate spontaneously, giving off radiations. The period of time required for half of the material to decompose in this way is called the 'half-life' of the element.

One of the most valuable of these radioactive isotopes for biological work is radioactive phosphorus ^{32}P. It emits β particles (electrons) and has a half-life of 14·5 days, that is half of it is decomposed in this period. It is usually prepared by the action on sulphur of neutrons in an atomic pile.

Radioactive phosphorus is usually employed in the form of sodium phosphate suitably diluted with sodium phosphate containing ordinary phosphorus ^{31}P. The mixture can be injected into animals and, after a suitable time, various phosphorus compounds can be isolated from the tissues and the concentration of ^{32}P in them determined. ^{32}P on decomposition gives off β-rays consisting of fast-moving electrons, by which the concentration of ^{32}P in the phosphorus compound can be determined with the aid of a *Geiger counter* or sensitive electrometer.

Radioactive carbon can be obtained in two forms, ^{11}C and ^{14}C (Fig. 19, 2).

^{11}C has a half-life of only 21 minutes and experiments carried out with it must be brief, but it has been used successfully for determining the fate of carbon monoxide in the body.

^{14}C has a half-life of 5568 years and is proving exceedingly useful in biological investigations. It is prepared by bombarding nitrogen atoms with fast neutrons:

$$^{14}N_7 + {}^{1}n_0 \rightarrow {}^{14}C_6 + {}^{1}H_1$$

a proton being ejected from the nucleus when the neutron enters. It decomposes spontaneously to yield nitrogen again, the nucleus of 6 protons and 8 neutrons giving place to one of 7 protons and 7 neutrons while an electron is ejected from the nucleus in the form of β-radiation:

$$^{14}C_6 \rightarrow {}^{14}N_7 + e$$

Radioactive isotopes of other elements have also been employed in biological work. They include sulphur (^{35}S), iron (^{59}Fe and ^{55}Fe), iodine (^{131}I), and the radioactive form of hydrogen, tritium ^{3}H. Two forms of radioactive sodium are known, ^{24}Na (half-life 14·8 hours) and ^{22}Na (half-life 3 years). They have proved useful in studies on the body fluids (Chapter 35).

REFERENCES

Francis, G. E., Mulligan, W. & Wormall, A. (1954). *Isotope Tracers.* London: Athlone Press.

'Isotopes in Medicine" (1952). *Brit. med. Bull.*, **8**.

CHAPTER 20

CARBOHYDRATE METABOLISM

THE main product of the digestion of carbohydrate in the alimentary tract is glucose, and it is therefore with the fate of glucose that we are chiefly concerned.

Small amounts of fructose and of galactose are formed as the result of the digestion of sucrose and of lactose respectively, but both these sugars are converted to glucose in the liver and may be built up into glycogen in the liver.

THE BLOOD SUGAR

Glucose is usually estimated by its power to reduce such reagents as cupric salts or potassium ferricyanide. Although the blood normally contains, in addition to glucose, small amounts of other reducing substances such as glutathione, these substances do not vary much in concentration in comparison with glucose. In the fasting human subject the concentration of 'true blood sugar,' that is glucose only, is about 80 mg. per 100 ml., although the total reducing power corresponds to about 100 mg. per 100 ml. With one of the most commonly used methods of estimation, that of Hagedorn and Jensen, the normal fasting blood-sugar level varies between 80 and 100 mg. per 100 ml. The concentration of glucose in human blood is the same in the cells as in the plasma.

After a meal containing carbohydrate the blood-sugar level rises (*hyperglycæmia*). This can readily be demonstrated in the human subject by means of blood-sugar estimations carried out during a standard procedure known as the *glucose-tolerance test*, widely used in the investigation of disease. The test is carried out as follows: A sample of blood is withdrawn from a vein or from a finger stab for the determination of the fasting blood-sugar level and urine is voided to empty the bladder. The subject, who should have had nothing to eat for at least eight hours, then drinks a solution of glucose containing 50 g. of glucose (or preferably one containing 0·75 g. per kg. body weight) dissolved in about 200 ml. water. Samples of blood and urine are taken at thirty-minute intervals over a period of two hours and the sugar concentrations estimated. The results of such a test in normal subjects are shown plotted in curves in Fig. 20, 1. The following points should be noted:

1. The blood sugar at the beginning of the test is at the fasting level of about 90 mg. per 100 ml.

2. The ingestion of glucose is followed by a rise in the blood-sugar level since the rate of absorption of glucose from the alimentary canal is greater than the rate at which the sugar is removed from the blood by conversion to glycogen and by oxidation.

3. The maximum level reached is about 140 to 150 mg. per 100 ml. and this level is reached at 30 to 45 minutes after glucose ingestion.

4. During the latter part of the test the blood-sugar level falls as the result of glycogen synthesis and carbohydrate oxidation and at the end of

two hours the blood-sugar level is back again to the normal value. In the course of its return to normal the blood-sugar level frequently falls below the initial fasting value (Fig. 20, 1).

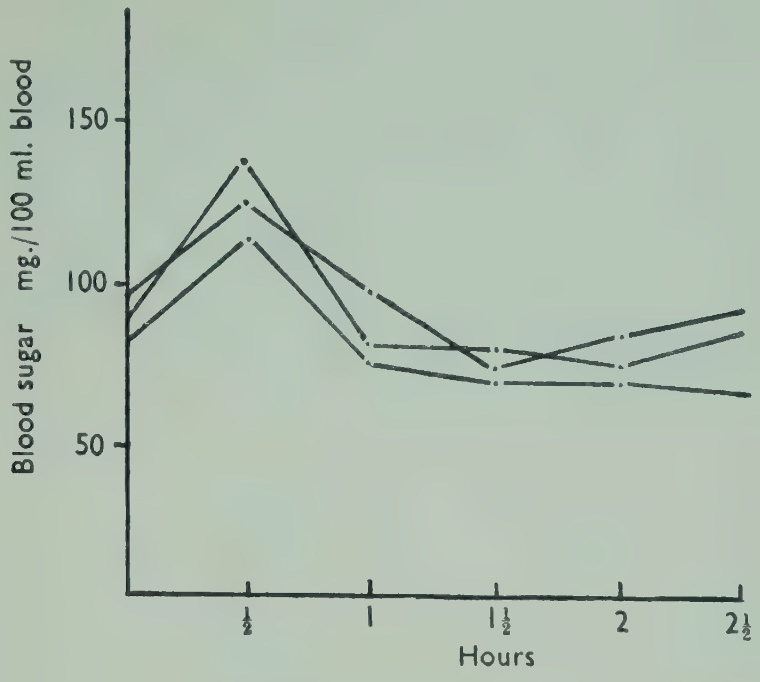


FIG. 20, 1

Blood-sugar curves from three normal human subjects after the ingestion of 50 g. of glucose (glucose tolerance test).

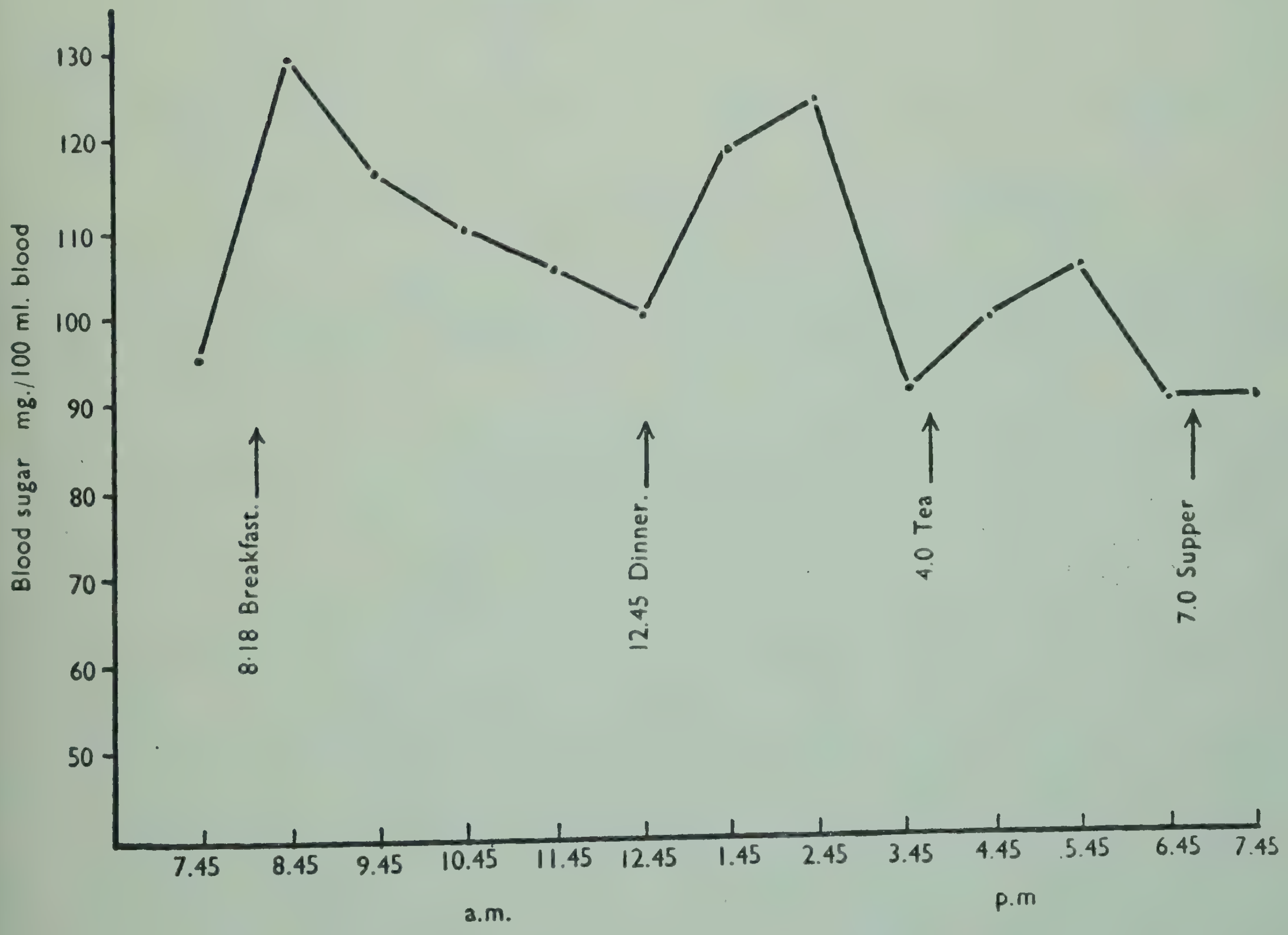


FIG. 20, 2

Capillary blood-sugar concentration during twelve hours, showing the effect of meals in a healthy human subject.

5. During the course of the test on normal persons glucose does not appear in the urine since the ability of the renal tubular cells to reabsorb glucose (Chapter 34) is not exceeded.

A rise and fall in blood-sugar level also follows the ingestion of a meal containing carbohydrate in any form, the height of the curve varying with the amount of glucose produced as the result of carbohydrate digestion, with the rate of digestion and with the rate of removal of glucose from the circulating blood (Fig. 20, 2).

In the fasting subject the sugar level is 2 to 3 mg. per 100 ml. higher in the arterial than in venous blood because sugar is being taken up by the tissues from the blood. The concentration of glucose in freely flowing capillary blood, such as may be obtained by stabbing the finger, is virtually the same as that found in true arterial blood. During the absorption of a meal containing carbohydrate the difference between venous and arterial blood sugar is considerably increased (Fig. 20, 3).

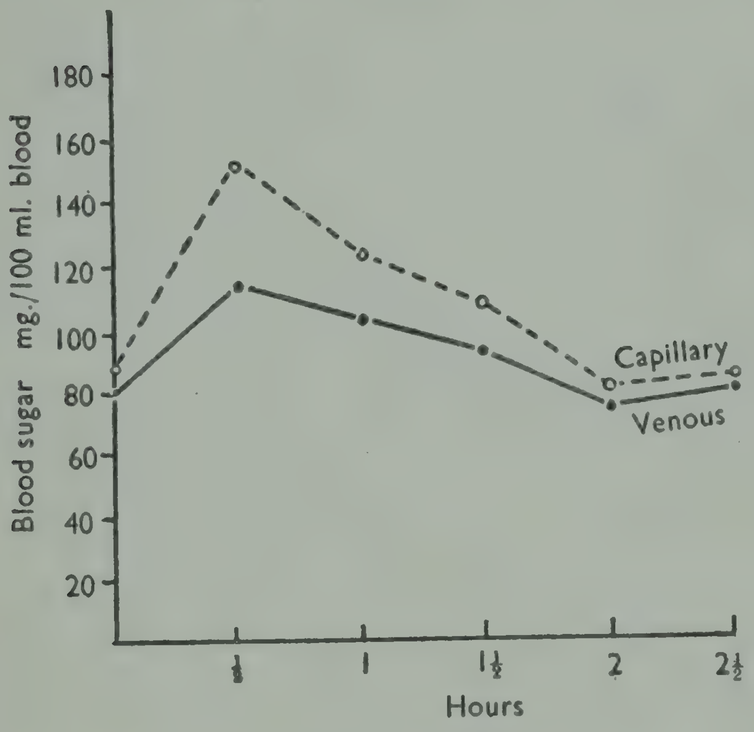


FIG. 20, 3

Capillary and venous blood-sugar levels in a healthy human subject after ingestion of 0·75 g. glucose per kg. body weight.

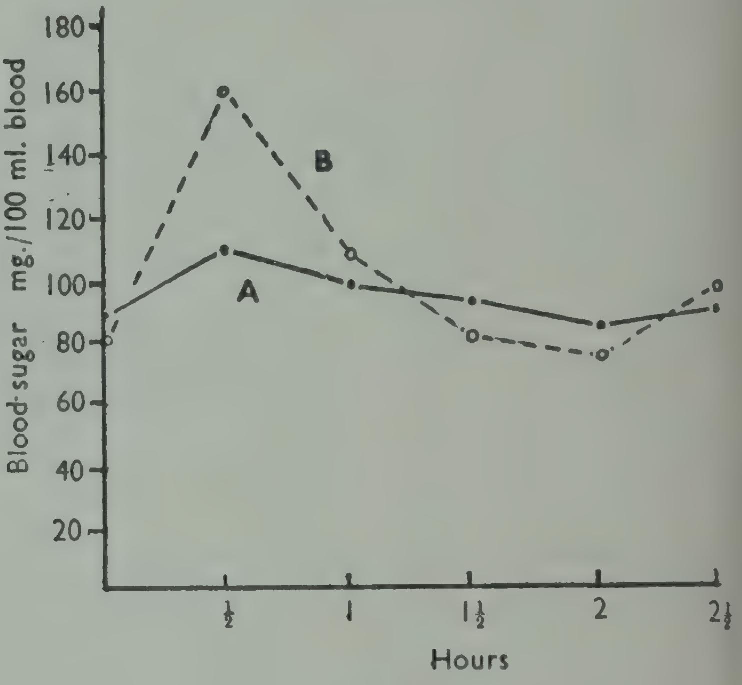


FIG. 20, 4

Blood-sugar curves from the same healthy human subject on (A) a high-carbohydrate low-fat diet and (B) a high-fat low-carbohydrate diet.

The type of curve given by normal subjects in the glucose tolerance test depends to some extent on the kind of food eaten during the few days previous to the test. If the diet for some days previously has contained little carbohydrate and much fat, glucose tolerance is low, that is the rise in blood sugar after ingestion of 50 g. of glucose is great, whereas when the diet has contained a high proportion of carbohydrate and little fat, glucose tolerance is high, a smaller rise in blood sugar following the ingestion of 50 g. of glucose (Fig. 20, 4). In persons with normal carbohydrate metabolism successive doses of glucose produce no further increase in blood sugar (Staub-Traugott phenomenon), whereas an impaired carbohydrate tolerance is revealed by a continuous increase in blood sugar following successive doses of 50 g. glucose. This fact is sometimes made use of in studying carbohydrate metabolism and is illustrated in Fig. 20, 5.

The extra glucose which appears in the portal blood may be disposed of in a number of different ways (Fig. 20, 6):

1. It may be stored as glycogen.
2. It may be converted to fat.
3. It may be oxidized in the tissues.

1. **Storage as Glycogen.**—Both the liver and the muscles normally contain large amounts of glycogen. A well-fed man weighing 70 kg., for example, has about 100 g. of glycogen in his liver and 250 to 350 g. of glycogen in his muscles.

After a meal containing carbohydrate has been eaten the glycogen content of the liver and of the muscles is increased. In the muscles this increase is due to glycogen synthesis from glucose (*glycogenesis*). The rise in the liver glycogen is due more to retardation of glycogen breakdown (*glycogenolysis*) (p. 293) than to glycogen synthesis.

The formation of glycogen from glucose is facilitated by insulin. If insulin is administered along with glucose the amount of glycogen formed in the muscles is much greater than when glucose is administered alone. This effect is much less marked when the blood-sugar level is very high, for example 400 mg. per 100 ml. (p. 303).

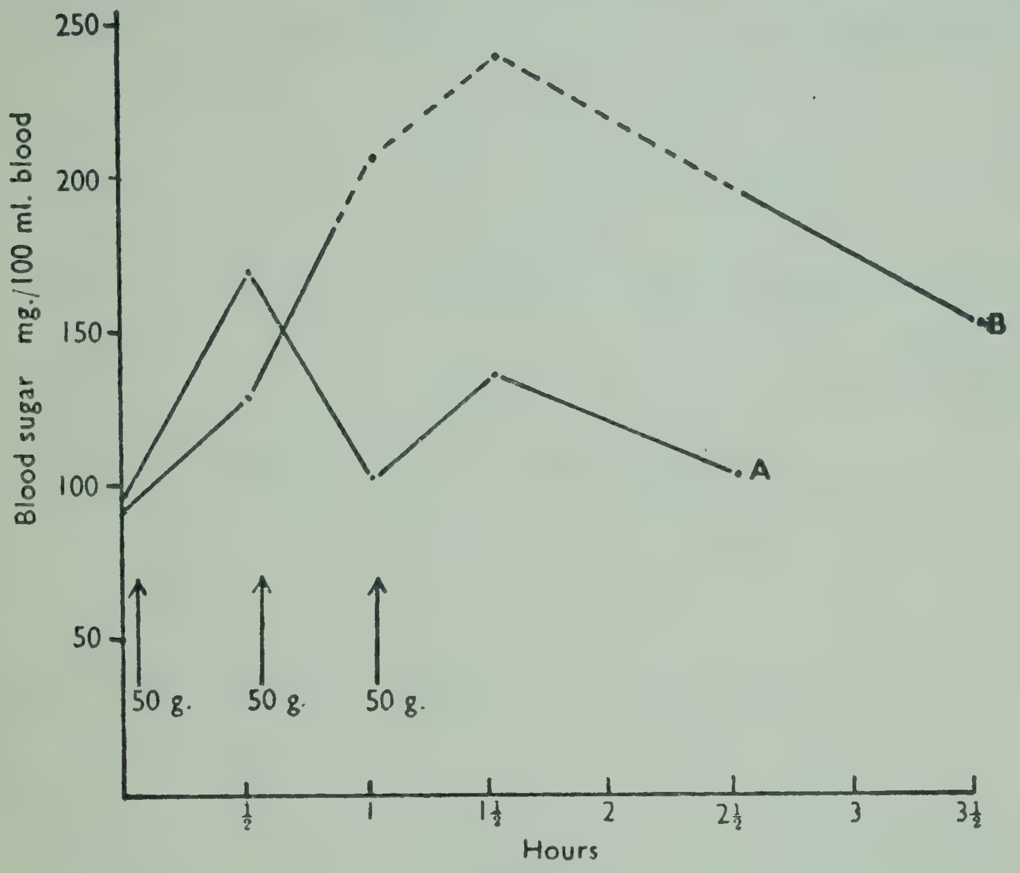


FIG. 20, 5

Response of blood sugar to successive doses of 50 g. glucose (Staub-Traugott phenomenon). The curves represent blood-sugar values at intervals after the ingestion of three doses of 50 g. glucose given at half-hourly intervals. Curve A is from a normal individual; B from a patient with impaired carbohydrate tolerance. Glycosuria is indicated by interrupted line.

2. **Conversion of Glucose to Fat.**—It is common knowledge that when human beings or animals eat excessive amounts of carbohydrate in any form they tend to put on fat. The classical experiments of Lawes and Gilbert at the Rothamsted Experimental Station more than eighty years ago demonstrated that the carcasses of well-fed pigs contained more fat than could be derived from the fat and protein which the animals had eaten during their lifetime. Some of the carcass fat must therefore have been synthesized from the carbohydrate of the food.

The exact steps by which sugars are converted into fatty acids are not yet known. It appears probable, however, that the carbohydrate molecules are broken down by way of pyruvic acid to simple units such as acetylcoenzyme A which is discussed below, and that several of these units then condense to form long fatty chains (see also p. 330). For the conversion, thiamine (vitamin B_1) diphosphate is essential, together with pyridoxine and pantothenic acid. Rats fed on a diet containing no thiamine are unable to synthesize fat from carbohydrate and protein.

3. **Oxidation of Glucose.**—One of the most important tissues in which carbohydrate is utilized is muscle, where glycogen breaks down anaerobically to pyruvic acid with the liberation of energy for muscular contraction.

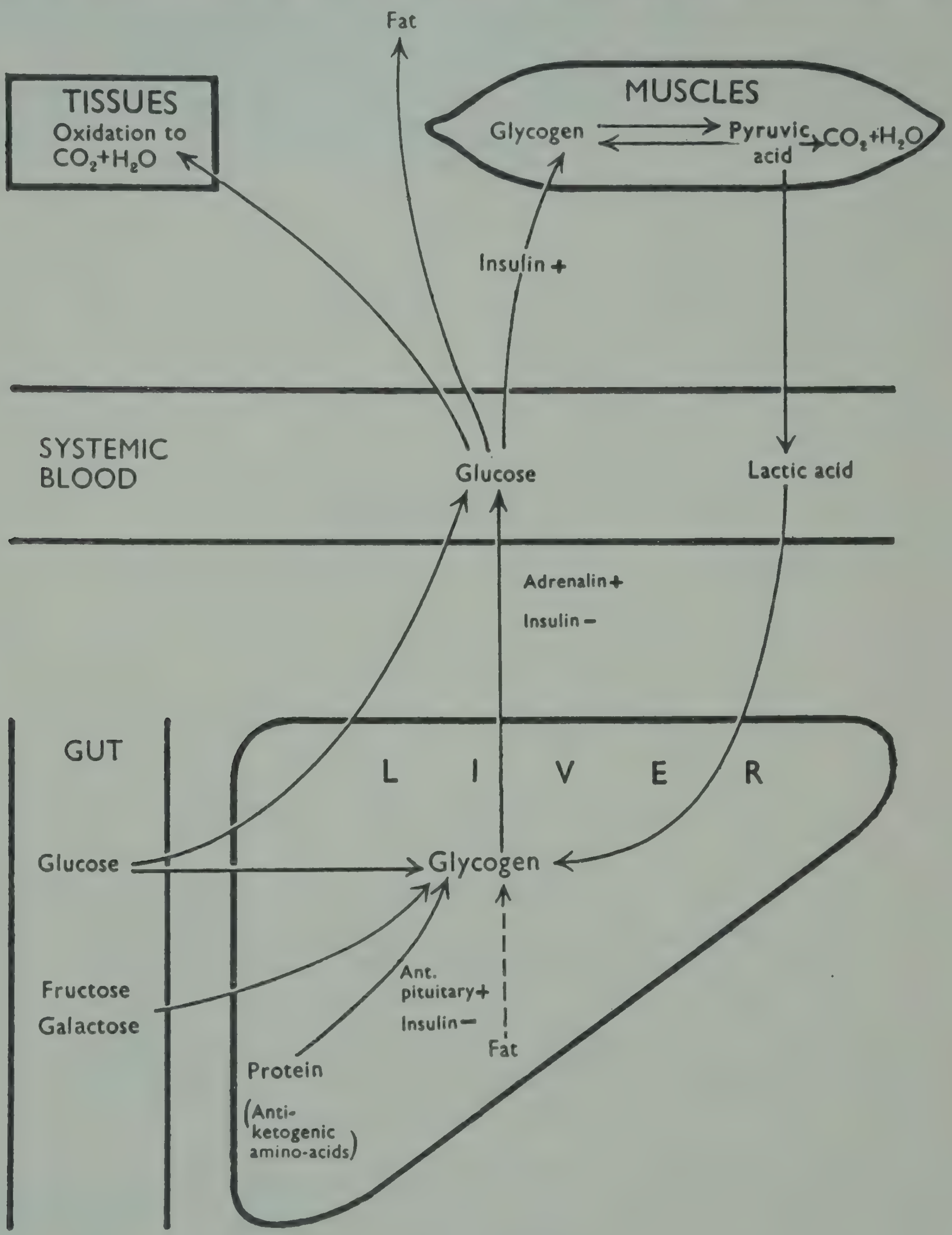

Fat
TISSUES
Oxidation to
CO_2+H_2O
MUSCLES
Glycogen
Pyruvic acid
CO_2+H_2O
Insulin +
SYSTEMIC BLOOD
Glucose
Lactic acid
Adrenalin +
Insulin −
GUT
L I V E R
Glucose
Glycogen
Fructose
Galactose
Ant. pituitary +
Insulin −
Fat
Protein
(Anti-ketogenic amino-acids)

Fig. 20, 6

Diagrammatic representation of carbohydrate metabolism.

In other tissues glucose from the blood may be utilized directly by oxidation to carbon dioxide and water. Nervous tissue, for example, apparently obtains its energy by direct oxidation of the blood glucose and, having no reserve of glycogen, is dependent upon an adequate level of circulating blood glucose. When the blood sugar falls to a low level (some 50 mg./100 ml. in a normal person), mental excitement and restlessness readily occur, followed, if the blood sugar falls much below 40 mg./100 ml., by muscular twitchings, increasing stupor and convulsions. In the treatment of certain nervous disorders (schizophrenia) insulin is injected in large enough amounts to produce prolonged coma but how this produces a beneficial effect is not at present clear.

The remainder of this chapter is concerned with the synthesis and breakdown of glycogen. The oxidation of glucose and its conversion into fat are discussed on pp. 316 and 328 respectively.

THE LIVER GLYCOGEN

Glycogen occurs in the liver as intracellular particles or storage granules which can be demonstrated by special staining reactions. In such a form it is protected to some extent from the metabolic activities of the surrounding protoplasm by the colloidal nature of the cytoplasm of the liver cell. Any injury to the cell renders the glycogen particles more vulnerable to attack. Hence adequate glycogen stores are associated with healthy tissue; and interference with the hepatic cells by disease or injury (infection, anoxia, or poisons) results in a fall in the glycogen content of the liver. Conversely, an adequate store of glycogen in the liver is essential for the protection of the protein structure. The likelihood of damage to the livers of experimental animals by various injurious agents, chloroform for example, is reduced if the organ is well stocked with glycogen, and therefore the treatment of liver disease in man involves the administration of adequate amounts of glucose.

It can easily be demonstrated by experiments on animals that the glycogen content of the liver is increased after the ingestion of carbohydrate, and the original conclusion of Claude Bernard in 1850 was that ingested glucose was converted into glycogen in the liver. Recent accurate measurements of the amounts of glucose entering and leaving the liver and other tissues have shown that this view is not quite correct. It is now considered that much of the glucose absorbed from the gut during digestion passes through the liver into the systemic circulation and is oxidized in the tissues or is converted into glycogen in the muscles. A relatively small amount is converted immediately to liver glycogen the bulk of which is formed from non-carbohydrate sources, or from carbohydrate sources other than glucose. Thus liver glycogen is formed from (1) certain amino-acids (see p. 345), (2) fructose and galactose, (3) lactic acid, which is formed in the muscles by breakdown of glycogen and may pass to the liver in the blood stream and there be converted into glycogen, and (4) fat, although the evidence that glycogen is formed in the liver from fat is inconclusive.

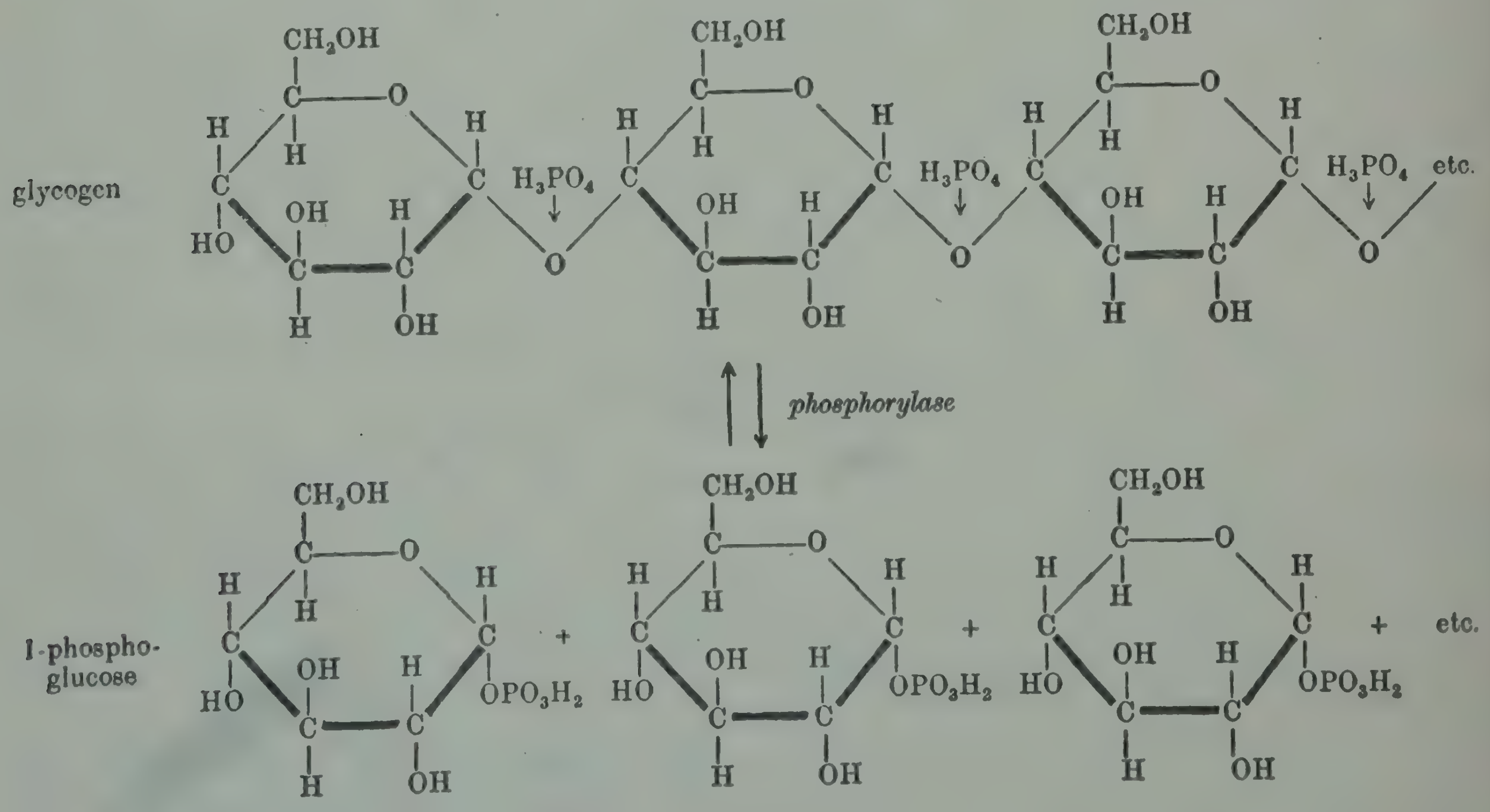
The liver glycogen synthesized from these various sources is continuously being broken down into glucose by the process of *glycogenolysis*. The liver then plays two important roles in carbohydrate metabolism. By breaking down glycogen it can liberate glucose into the circulating blood to maintain the blood sugar at its normal level. The second or glycogenic function is the synthesis of liver glycogen from glucose or other precursors.

In animals from which the liver has been removed the blood-sugar level quickly falls and death results very soon unless glucose is continuously injected into the blood stream. The tissue glycogen is not directly available as a source of glucose and the glycogen stores in the muscles are unable to contribute to the blood sugar.

GLYCOGENOLYSIS IN THE LIVER.—The term *glycogenolysis* indicates the breakdown of glycogen. In the liver the product of glycogenolysis is glucose which passes out into the blood as blood sugar. In muscle glycogen breaks down to pyruvic and lactic acid, and this process is usually referred to as *glycolysis*. It will be discussed in detail later (p. 311).

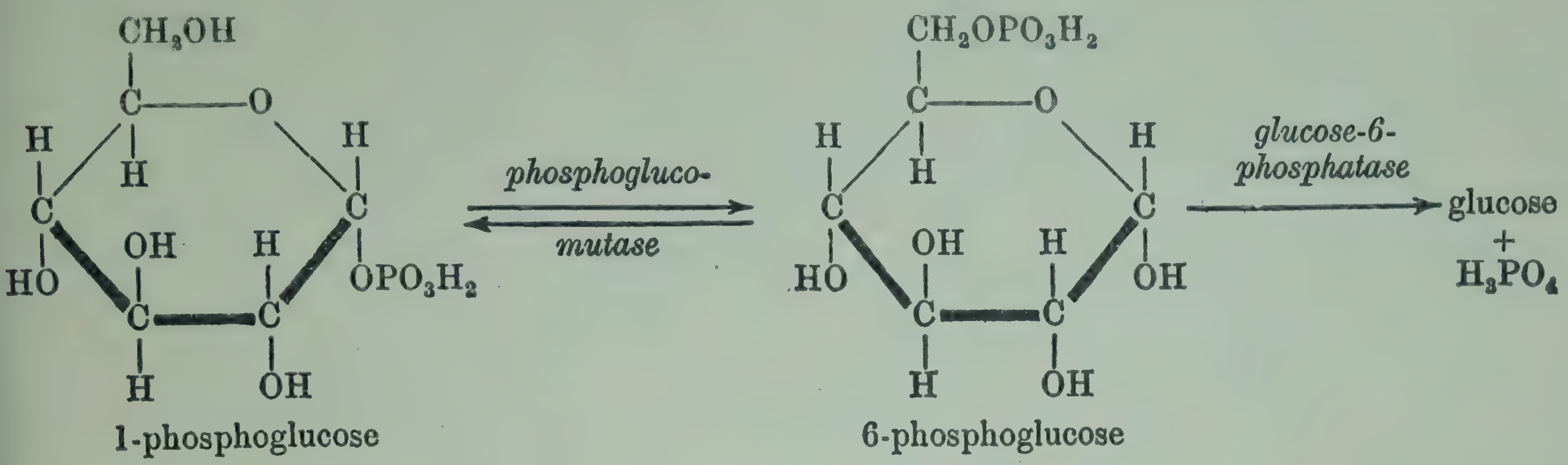
Starch and glycogen can, of course, be broken down by direct hydrolysis under the influence of such enzymes as diastase (amylase), but diastase is not found to any appreciable extent in the liver and it does not appear to play any part in the normal breakdown of liver glycogen. While hydrolysis is a normal process during digestion in the gut, and although diastase can be demonstrated in the blood, it is unlikely that it takes place to any important extent in the cells of tissues under normal conditions. There is, however, evidence that anoxia and infections in animals may be associated with a significant degree of diastase activity within the liver cells. This may be one of the reasons for the lower levels of liver glycogen often found in these conditions, and even for the well-known fact that diabetic patients with acute infections are remarkably resistant to insulin. We shall see (p. 301) that insulin is concerned with certain processes involving phosphorylation; it has no effect on the action of diastase. The breakdown of glycogen by diastase will therefore continue independently of the amount of insulin injected. Diastase presumably from the salivary glands and the pancreas is normally found in small amounts in the blood and urine as an overflow from these glands.

In the liver, and also in muscle as we shall see later, glycogen is broken down not by direct hydrolysis but by an analogous process called *phosphorolysis*. Under the influence of the enzyme *phosphorylase*, and its coenzyme adenylic acid, the long glycogen chain is broken down by the introduction of phosphate groupings at the linkages between the individual glucose residues.



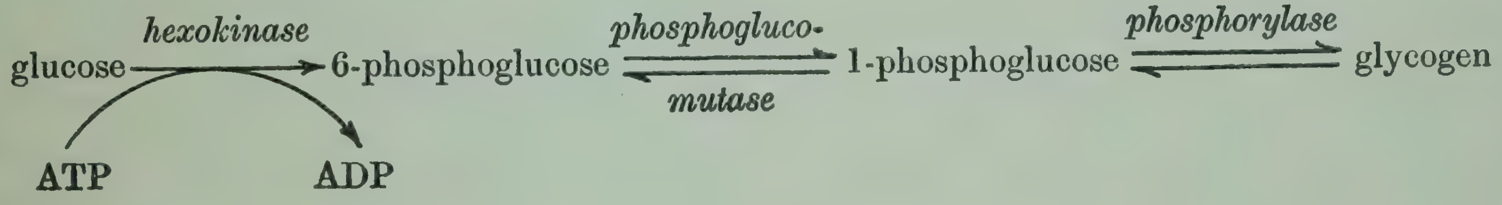
Whereas in hydrolysis the elements of water are introduced at the broken linkages, in phosphorolysis phosphoric acid is introduced instead. The process is catalysed by the enzyme *phosphorylase* and results in the formation of 1-phosphoglucose which is converted by the enzyme *phosphoglucomutase* to 6-phosphoglucose. Liver, unlike muscle, contains an active phosphatase (*glucose-6-phosphatase*) which breaks down

6-phosphoglucose to phosphoric acid and glucose. Thus the blood sugar is formed from liver glycogen by way of the phosphoglucoses (Fig. 20, 7).


phosphogluco-
mutase
glucose-6-
phosphatase
glucose
+
H_3PO_4
1-phosphoglucose
6-phosphoglucose

This reaction is accelerated by adrenaline which facilitates phosphorylation of glycogen. The accelerating effect is not obtained in isolated enzyme systems but is found only in the intact cell. It is probable that adrenaline acts by increasing the concentration of active phosphorylase enzyme in the cell.

Glycogenesis.—During the process of glycogen synthesis (*glycogenesis*) in liver and muscle, glucose is phosphorylated under the influence of the enzyme *hexokinase* in the presence of the phosphate donor adenosinetriphosphate (ATP) (p. 61). The ester formed is 6-phosphoglucose which is converted into 1-phosphoglucose under the influence of phosphoglucomutase. The phosphorylase reaction is then reversed and glycogen is produced.


hexokinase
glucose
6-phosphoglucose
phosphogluco-
mutase
1-phosphoglucose
phosphorylase
glycogen
ATP
ADP

Glycogen has, in fact, been prepared from glucose *in vitro* by making up a system consisting of the three enzymes hexokinase, phosphoglucomutase and phosphorylase, together with adenosine triphosphate. This system converts glucose successively to 6-phosphoglucose, 1-phospho-

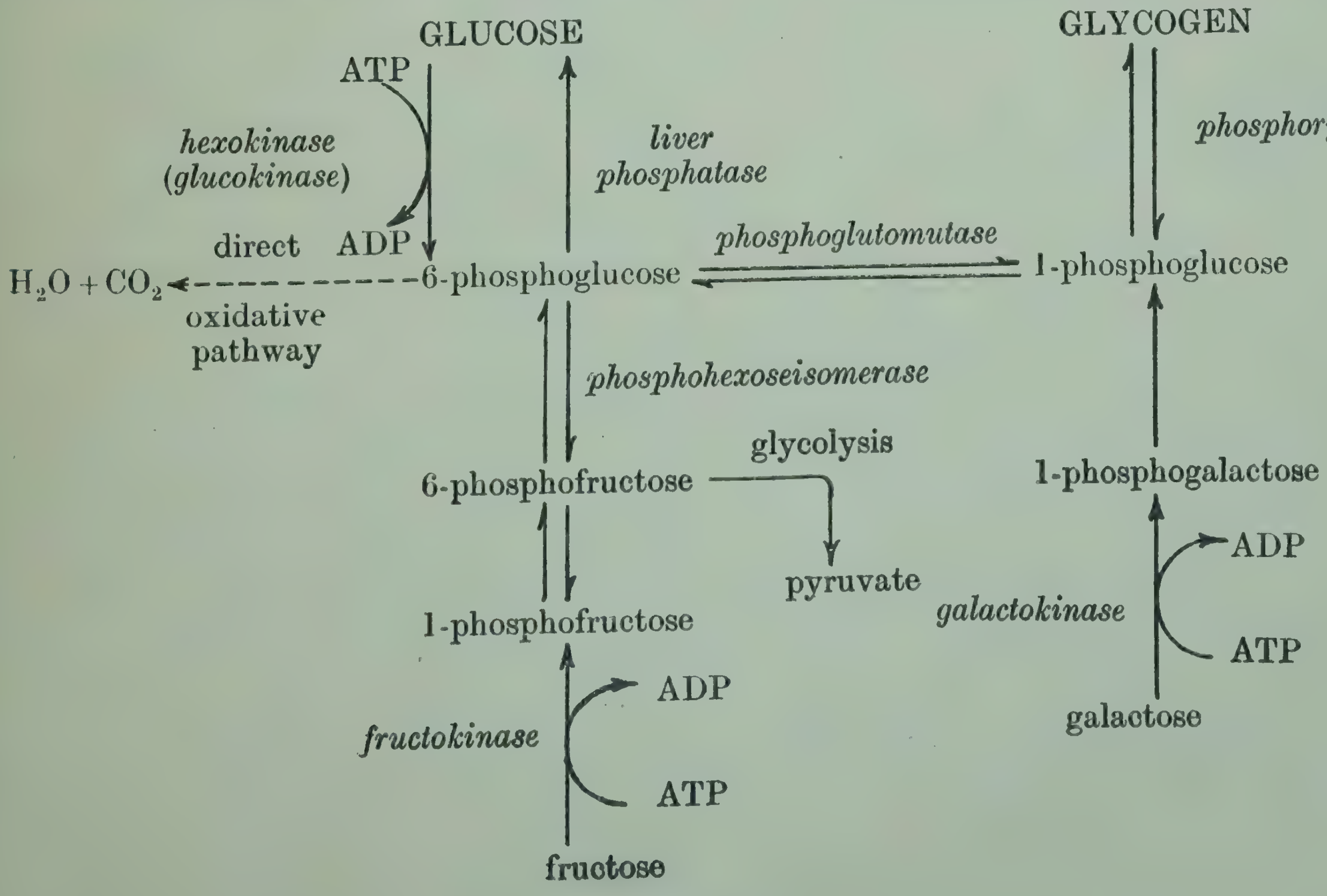

GLUCOSE
GLYCOGEN
ATP
hexokinase
(glucokinase)
liver
phosphatase
phosphorylase
direct
ADP
phosphoglutomutase
$H_2O + CO_2$
6-phosphoglucose
1-phosphoglucose
oxidative
pathway
phosphohexoseisomerase
glycolysis
6-phosphofructose
1-phosphogalactose
pyruvate
ADP
galactokinase
ATP
1-phosphofructose
ADP
fructokinase
ATP
galactose
fructose

Fig. 20, 7

glucose and glycogen. The glycogen prepared artificially by this process gives a brown colour with iodine (p. 17) and appears to be identical with the naturally occurring polysaccharide. It has been observed that the enzyme hexokinase is inhibited by anterior pituitary hormone preparations and that this inhibition is abolished by insulin.

The synthesis of glycogen *in vitro* in this way is not affected by the addition of insulin, and even *in vivo* small amounts of glycogen may be formed from glucose in the absence of insulin. Glycogen may be deposited in the liver and muscles of depancreatized dogs, but only when the blood-sugar level is very high (p. 304). At normal blood-sugar levels in healthy subjects, with adequate supplies of sugar, insulin facilitates the deposition of glycogen in the muscles and to a smaller extent perhaps in the liver.

It is well known that the ingestion of fructose increases the liver glycogen. The first step in this conversion appears to be the phosphorylation of fructose to give phosphofructose, which is then converted to 6-phosphoglucose and so by way of 1-phosphoglucose to glycogen (Fig. 20, 7). The phosphorylation of fructose requires oxidative energy and occurs in tissue slices only if sufficient oxygen is present. By a similar series of reactions fructose may also be converted to glucose, the 6-phosphoglucose being attacked by phosphatase which splits it into glucose and inorganic phosphate. Because the conversion of fructose to glucose is confined to the liver and is independent of insulin, the concentration of fructose in the blood following the ingestion of 50 g. of the sugar has been used as an index of hepatic function. In liver disease the hepatic cells are unable to convert fructose to glycogen at the normal rate and the concentration of fructose in the systemic blood rises above the normal. Severe disease of the liver also interferes with the storage of glycogen, with the result that hyperglycæmia and glycosuria follow the ingestion of 50 g. of glucose (Fig. 20, 8).

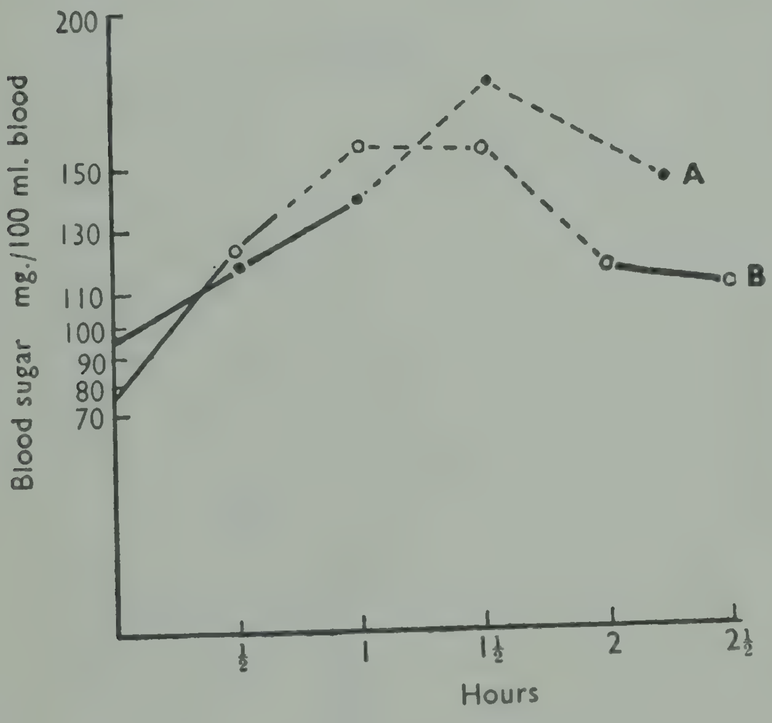


FIG. 20, 8

Blood-sugar curves (capillary blood) after oral administration of glucose 0·75 g./kg. in two patients, A and B, with liver disease. Interrupted line = periods of glycosuria. Although of the 'diabetic type' (Fig. 20, 14) these curves have a characteristically low fasting level.

Galactose can also give rise to glycogen in the liver. This conversion occurs by way of the phosphorylated derivative, 1-phosphogalactose, formed under the influence of the enzyme galactokinase. The conversion involves the participation of the coenzyme uridine diphosphate glucose (UDPG).

GLUCONEOGENESIS. THE GLYCOGENIC FUNCTION OF THE LIVER.—We have seen how glycogen may be built up from glucose by way of the phosphorylated derivatives. Fructose and galactose also can be converted in the liver into glucose, and can act as sources of liver glycogen. In addition liver glycogen can be built up from non-carbohydrate precursors which are converted first into glucose and then into glycogen—a process which is known as *gluconeogenesis*. We shall see, for example (p. 345), that certain of the amino-acids are antiketogenetic, that is to say, these amino-acids after deamination yield keto acids which can be built up into liver glycogen. Since the stores of glycogen in the liver are sufficient to last the body for only a short time, the process of liver glycogen formation from the antiketogenic amino-acids of protein is an important one, and a large part of protein ingested is in fact utilized for the synthesis of liver glycogen. During starvation tissue protein is actually broken down and its antiketogenic amino-acids used for conversion to liver glycogen to maintain the blood-sugar level.

Another possible source of liver glycogen is fat. The problem of the conversion of fat to carbohydrate has given rise to much controversial argument over many years and is not yet finally settled. The evidence cited in favour of such a conversion is largely indirect and cannot be discussed in detail here. It may be said, however, that in recent years fairly convincing evidence has been put forward in support of the view that glucose may be formed from fat in very small amounts.

The formation of liver glycogen from non-carbohydrate precursors is under the control of the endocrine system (p. 300). While the hormones of the anterior lobe of the pituitary gland, of the adrenal cortex and of the thyroid gland can all increase the rate of gluconeogenesis in the liver, they are not effective in the absence of a normal supply of insulin.

If glycogenesis be impaired for any reason, the liver may no longer be able to supply a sufficient amount of glucose to the blood. Mild hypoglycæmia will then result during periods when the body is largely dependent on the liver as the source of blood sugar, as for example after a prolonged fast. Because hypoglycæmia is often attended by the feeling of hunger, and by 'hunger-pains' in the stomach, it has been suggested that this is one of the mechanisms for regulating the intake of food (p. 215). As the glycogenic process is delicate and easily impaired, mild hypoglycæmia before meals may occur in certain forms of liver disease. It may also occur during general anæsthesia. From these considerations, and because it has been shown (p. 293) that a liver depleted of glycogen is more susceptible to injury, it will be realized how undesirable is a prolonged fast immediately before a surgical operation.

The Origin of the Blood Sugar

We are now in a better position to understand how the blood sugar is derived from various sources.

The main function of the liver in carbohydrate metabolism is the building up and breaking down of glycogen rather than its storage. Under fasting conditions, when sugar is not reaching the blood from the alimentary canal, the rate of liver glycogenolysis is equal to the rate of blood-sugar utilization in the extra-hepatic tissues and the blood-sugar level remains relatively constant. At the same time gluconeogenesis is proceeding in the liver from non-carbohydrate materials and the hepatic glycogen stores are maintained.

These stores are never great. Soskin has estimated that a man weighing 70 kg., with a liver weighing 1800 g., has a total liver glycogen store of about 100 g. If such a man were fasting and if gluconeogenesis were to cease, this amount of glycogen could provide the carbohydrate requirements of his extra-hepatic tissues (about 0·25 g. per kg. per hour) for only six hours. The liver glycogen thus represents a reserve of carbohydrate from which the liver may draw quickly while the process of gluconeogenesis is developing. When sugar is being absorbed from the intestine the rate of secretion of glucose by the liver into the blood stream is reduced. The process of gluconeogenesis in the liver continues and glycogen accumulates in the liver. When sugar is not available from the gastro-intestinal tract the process of gluconeogenesis provides the liver glycogen from which the blood sugar is supplied (Fig. 20, 9). In the latter event glycogenolysis provides glucose which is carried by the blood stream to the other tissues of the body where it can be oxidized with the liberation of energy. In the muscles the blood glucose is built up into glycogen which yields lactic and pyruvic acids during the process of muscular contraction. Some of the lactic acid formed from glycogen in the muscles in moderate or severe exercise diffuses out into the blood stream and is carried back to the liver where it is built up again into glycogen.

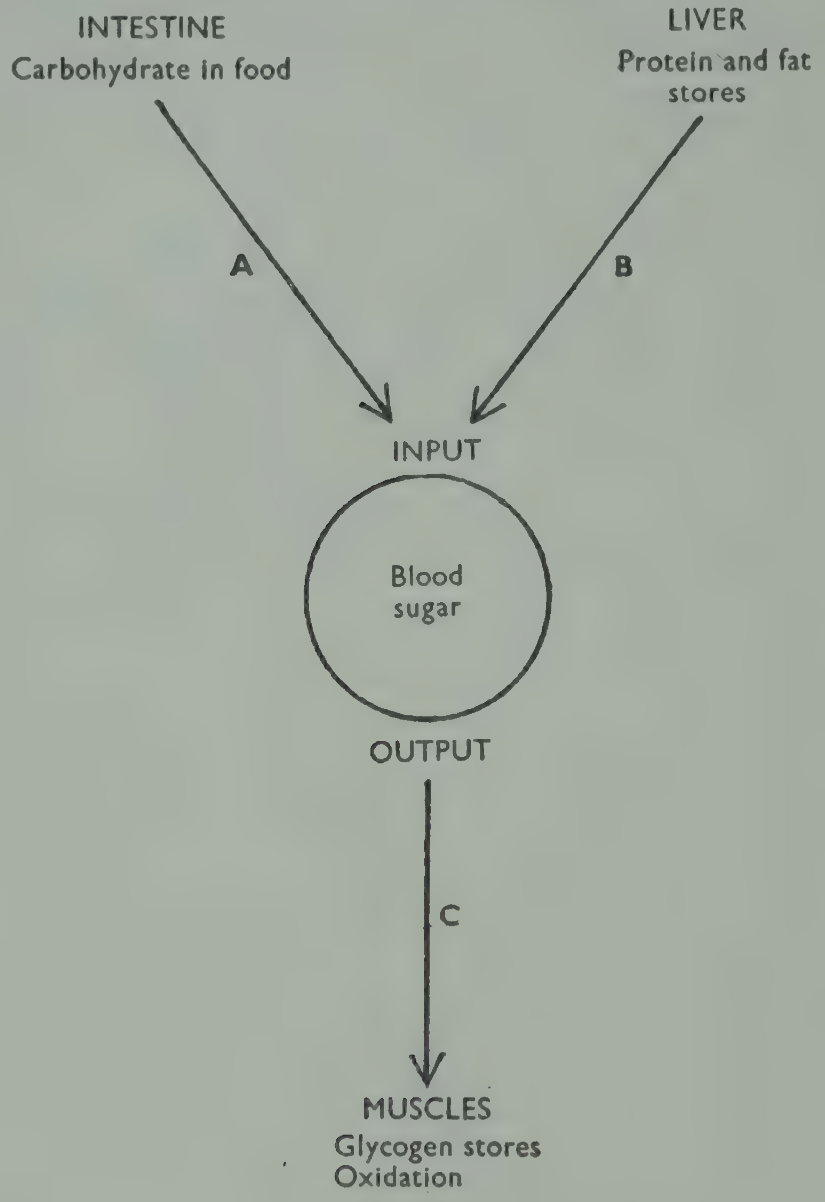


FIG. 20, 9

Effect of various factors on the blood-sugar level. Insulin increases C and decreases B. A meal increases A. Exercise increases C. Anterior pituitary increases B.

The blood normally contains small amounts of lactic acid (5 to 20 mg. per 100 ml.) which is believed to be formed from the glucose in the red cells. In conditions of severe exercise lactic acid diffuses out of the muscles into the blood and the blood lactic acid may rise to high levels, for example

200 mg. per 100 ml. After an injection of adrenaline the lactic acid level of the blood is raised appreciably as the result of increased muscle glycolysis.

The processes which maintain the circulating blood sugar at normal levels are excellent examples of *homoiostatic mechanisms.* The working of this regulating mechanism, in so far as the liver is concerned, has been demonstrated by Soskin and his collaborators in experimental animals by measuring the rate of blood flow through the liver and the sugar concentration in the blood. The amounts of sugar entering and leaving the liver per unit of time could then be calculated. In fasting animals the liver was found to be liberating sugar into the blood at a constant rate. When glucose was administered intravenously to the animal the blood-sugar level rose, and the liver stopped putting out sugar and started to take it in in large amounts. From the second to the third hour after glucose administration the liver neither took in nor put out sugar and the blood sugar fell slightly below the original level. After the period of reduced output the liver again began to supply sugar to the blood at the original rate.

An experiment of this sort demonstrates conclusively that one function of the liver is to act as a reservoir from which the blood-sugar level may be maintained while sugar is not entering the blood from other sources. If a supply from other sources is available, the supply from the liver is temporarily cut off.

The conversion of glucose to glycogen may be seriously interfered with in certain disorders of the liver characterized either by functional disorganisation of liver tissue (hepatitis) or by accumulation in the liver cells of excessive amounts of fat or glycogen (glycogen disease). In such circumstances the increase in blood-sugar level after glucose ingestion is not followed by glycogen synthesis and hyperglycæmia and glycosuria occur. Conversely the blood sugar falls to low levels after fasting.

Carbohydrate Metabolism Summarized

The steps in carbohydrate metabolism so far discussed are summarized in Fig. 20, 7 from which it will be noted that 6-phosphoglucose occupies a key position. It is formed both from the blood sugar by the hexokinase reaction and from glycogen, reversibly, by way of 1-phosphoglucose. It is an intermediate in the formation of glycogen from both glucose and fructose.

6-Phosphoglucose can be oxidized directly to carbon dioxide and water but the extent to which this reaction occurs is not known. Glycolysis is the most important mechanism for the breakdown of carbohydrate. This process, which occurs chiefly in muscle, is discussed in greater detail on p. 311. In glycolysis glycogen is broken down by way of 1-phosphoglucose and 6-phosphoglucose and a large number of intermediate compounds (Fig. 20, 17) to pyruvic acid ($CH_3 \cdot CO \cdot COOH$). In certain circumstances pyruvic acid may give rise to lactic acid but more usually it loses carbon dioxide to form a compound known as acetylcoenzyme A (a 'C_2-unit') which then enters what is known as the citric acid cycle (p. 316) for oxidation to carbon dioxide and water. This cycle is discussed in detail later, but it should be noted here that fatty acids on oxidation also yield C_2-units which are likewise oxidized by way of the citric acid cycle.

THE HORMONAL CONTROL OF THE BLOOD-SUGAR LEVEL AND OF CARBOHYDRATE METABOLISM

In a fasting man the blood-sugar level is maintained relatively constant largely through the opposing action of different endocrine secretions. The chief sources of hormones controlling carbohydrate metabolism are:

1. The anterior lobe of the pituitary gland.
2. Interalveolar cell islets (Islets of Langerhans) of the pancreas.
3. The medulla of the adrenal gland.
4. The cortex of the adrenal gland.

1. **The Pituitary Gland in Carbohydrate Metabolism.**—The connexion between the pituitary gland and carbohydrate metabolism was established by the classical experiments of Houssay and his collaborators in Buenos Aires from 1924 onwards. They showed that:

(*a*) Removal of the pituitary gland from a normal animal caused an increase in sensitivity to insulin, that is a given dose of insulin then produced a larger fall of blood sugar.

(*b*) Removal of the pituitary gland from an animal rendered diabetic by pancreatectomy was followed by a diminution in the intensity of the diabetic condition (Houssay dog).

(*c*) Injection of extracts of the anterior lobe of the pituitary gland into normal or hypophysectomized animals diminished their sensitivity to the hypoglycæmic action of insulin.

(*d*) Injection of such extracts into hypophysectomized and depancreatized dogs increased the severity of their diabetic state.

(*e*) Administration of extracts of the anterior pituitary gland to normal animals could induce diabetes.

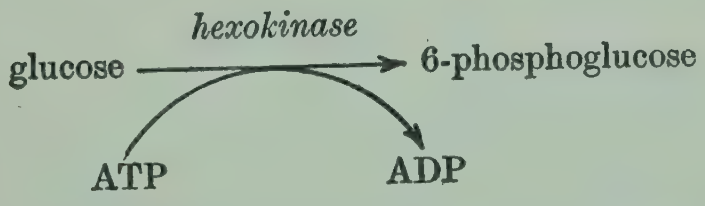
The importance of the last observation was emphasized by the work of F. G. Young in London. He found that when extracts of the anterior lobe were injected into dogs a latent period of three to five days elapsed before the blood sugar rose. During this latent period the animals became relatively insensitive to insulin, that is the same dose exerted a smaller effect in reducing the blood sugar. If the injections were continued beyond the latent period manifestations of diabetes mellitus began to appear, such as hyperglycæmia, glycosuria, ketonuria, polyuria and polydipsia, but the body weight did not fall and the liver glycogen levels were maintained. If injections of the same amounts of anterior pituitary extract were continued, the diabetic condition gradually disappeared, the blood-sugar level fell to the normal value, but the animals remained relatively insensitive to the hypoglycæmic action of insulin. On the other hand, if the amount of anterior pituitary extract injected daily was increased every few days the diabetic condition did not disappear, and continued indefinitely even after cessation of treatment with the extract. Permanent diabetes could thus be induced by administration of anterior pituitary extract over a period of a few weeks.

Glycosuria in such permanently diabetic dogs could be controlled by administration of insulin, but the animals required larger amounts of insulin than did depancreatized dogs under the same conditions. Insulin, however, was not essential for life since the dogs could survive for long periods without it despite heavy glycosuria, and did not even lose much weight provided they received sufficient food. If, however, the insulin was suddenly withdrawn, the animals died in a state resembling diabetic coma. In the permanently diabetic dogs degenerative changes could be found in the cells of the Islets of Langerhans.

While it is obvious that extracts of the anterior pituitary gland have a diabetogenic action, it is not certain that they contain a separate diabetogenic principle, and recent evidence suggests that the diabetogenic substance may be identical with the growth-promoting principle of the anterior pituitary gland (Chapter 53). Indeed crude pituitary extracts which are diabetogenic in the adult dog and cat cause enhanced growth in puppies and kittens. A suggestion that these effects might be due to separate factors in pituitary extracts has had to be abandoned as the result of recent successful efforts to purify the growth hormone and it is now clear that a single agent is responsible both for the development of diabetes in the adult animal, and for the stimulation of growth in the young animal. Indeed, if puppies are treated by injections of growth hormone, they grow up to the time when growth normally ceases and then, if the injections are continued, they become diabetic. On the other hand, rats receiving growth hormone continue to grow, even in old age, and never become diabetic. It is of interest to note that in the lactating cow growth hormone enhances the milk yield.

F. G. Young has therefore expressed the view that growth and diabetes induced by growth hormone are alternative responses to the same general metabolic stimulus. Administration of growth hormone to rats and dogs causes increased deposition of protein, increased oxidation of fat and decreased oxidation of glucose. In the growing animal this last effect results in deposition of carbohydrate in the tissues whereas in the adult it produces the hyperglycæmia and glycosuria of diabetes.

These results are concerned with the action of anterior pituitary extracts on carbohydrate metabolism in general but it is now possible to indicate at least one specific reaction in carbohydrate metabolism which is subject to pituitary influence. Cori has shown that the uptake of glucose by isolated muscle preparations is inhibited by anterior pituitary extract. The reaction involved here is the conversion of glucose to 6-phosphoglucose under the influence of the enzyme hexokinase.



While the hexokinase reaction is inhibited by anterior pituitary extract, insulin abolishes the inhibition and hence accelerates the reaction. It should be noted that when the system contains no anterior pituitary extract the addition of insulin does not accelerate the hexokinase reaction.

Recent work has demonstrated that the adrenocorticotrophic hormone (ACTH) of the pituitary gland has the effect of potentiating the action of the growth hormone in producing diabetes.

2. **The Interalveolar Cell Islets (Islets of Langerhans) of the Pancreas.** (*a*) **Insulin.**—The action of insulin on the blood-sugar level is the converse of that of the anterior lobe of the pituitary gland and of adrenaline. Normally the blood sugar is maintained relatively constant

under the opposing actions of insulin on the one hand and the anterior pituitary, adrenal cortex and adrenaline on the other. A high blood-sugar level such as follows adrenaline secretion invokes insulin secretion, with the result that glycogenolysis is inhibited and the normal blood-sugar level is restored.

When the normal supply of insulin to an animal is cut off by removal of the pancreas (*pancreatectomy*), or by damage to the secretory cells, the blood sugar rises and liver glycogen falls. Administration of insulin by injection reduces the blood sugar and stops the fall in liver glycogen.

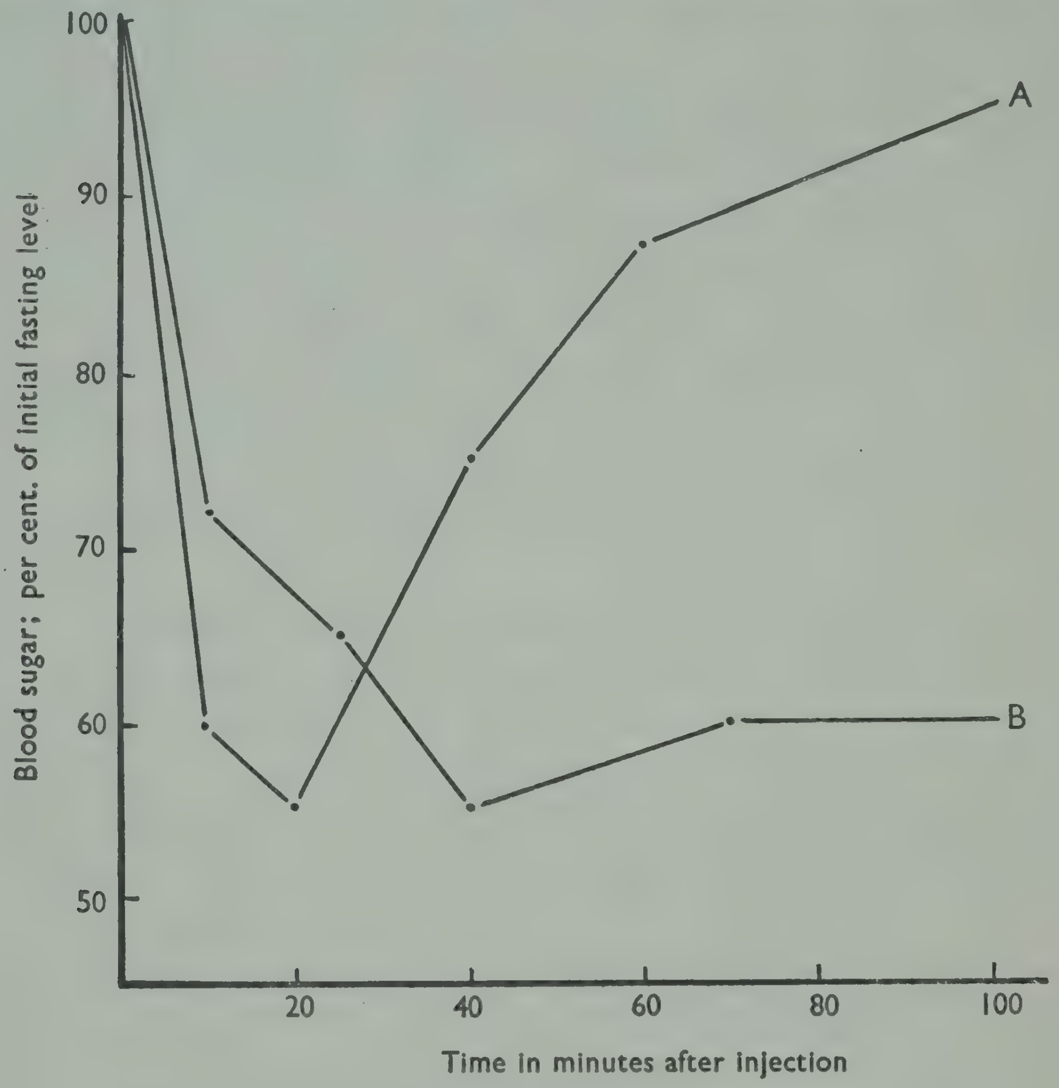


FIG. 20, 10

The alteration in concentration of blood sugar following intravenous injection of insulin (0·1 unit/kg.) in two patients. *A* had no disturbance of carbohydrate metabolism; *B* had a disorder of the adrenohypophysis (Simmond's disease) as the result of which gluconeogenesis was impaired and the blood sugar remained low.

When insulin is administered to a normal human subject the blood-sugar level falls. If it should fall below about 40 mg. per 100 ml. symptoms and signs characteristic of hypoglycæmia become evident, such as headache, difficulty in concentration, a sense of apprehension, tremor, sweating, convulsions and coma. The condition can be quickly remedied by giving glucose or cane sugar by mouth, by injecting glucose intravenously or, less certainly, by a subcutaneous injection of adrenaline. The effect of the injection of adrenaline is dependent upon there being sufficient glycogen in the body from which glucose may be produced. This condition is, however, not always fulfilled. Many of the symptoms of hypoglycæmia are due to the liberation of adrenaline from the adrenal glands in response to the low blood-sugar level rather than to the latter itself.

The behaviour of the blood-sugar level after the administration of a standard dose of insulin (0·1 unit per kg. body weight) is sometimes used as a diagnostic test (Fig. 20, 10) but the sensitivity of the body to the hypoglycæmic action of insulin varies with the diet. We have already seen how the sugar tolerance of the human subject increases when the diet in the few days preceding the test has contained a high proportion of carbohydrate and a low proportion of fat (Fig. 20, 4). In the same way it may be shown that, when a fasting human subject who has subsisted on such a diet is given a standard dose (3 units) of crystalline insulin intravenously, the blood sugar, after a very short latent period of about two minutes, quickly drops to below 70 mg. per 100 ml. If the subject has been subsisting on a low-carbohydrate high-fat diet during the few days preceding the test, the latent period before insulin begins to act is longer (7 min) and the blood-sugar level falls only to about 85 mg. per 100 ml. in the course of fifteen minutes (Fig. 20, 11). Sensitivity to insulin is therefore greater when the proportion of carbohydrate in the diet is high and the proportion of fat low (see also p. 290).

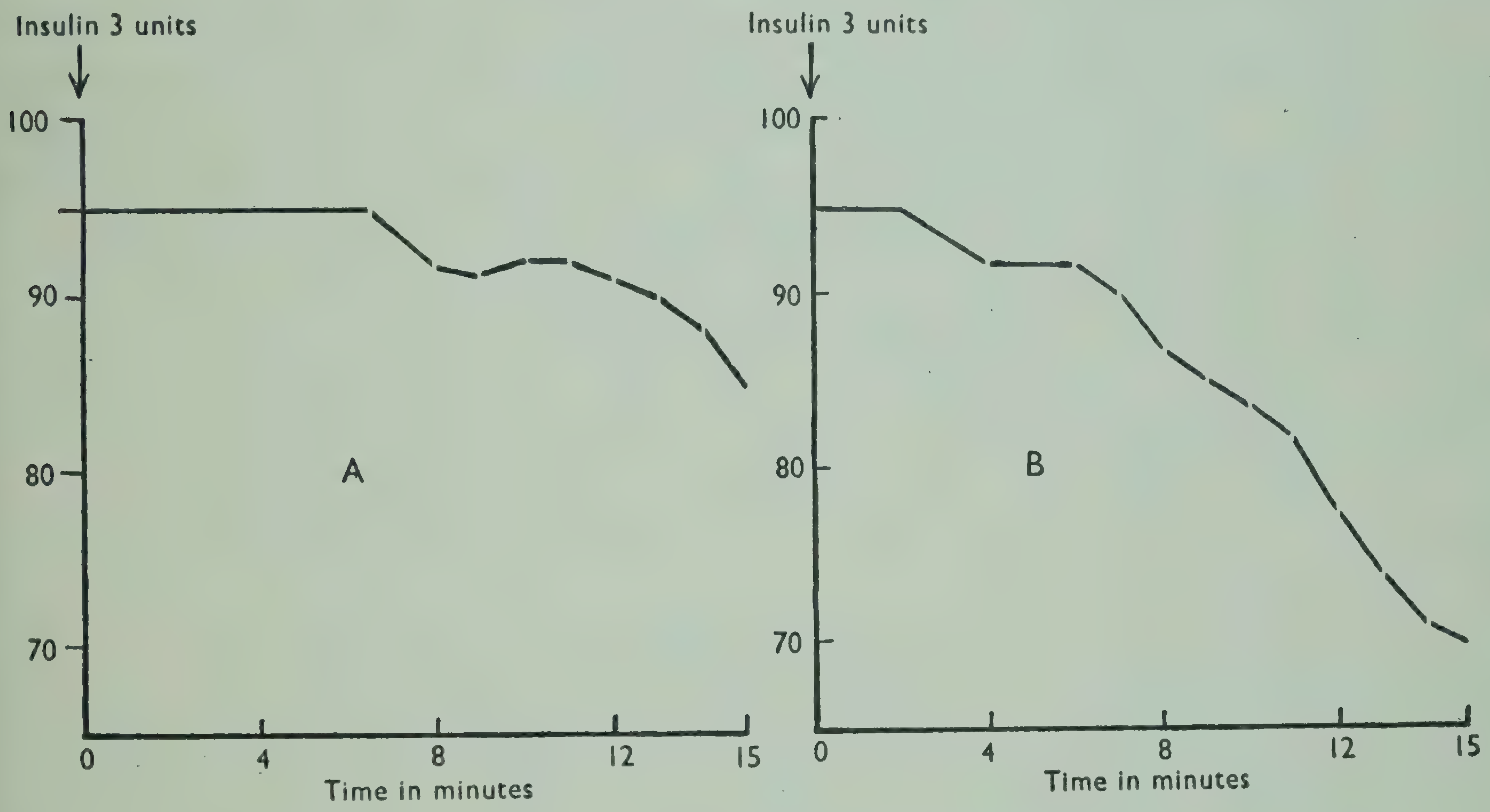


Fig. 20, 11

Effect of an intravenous injection of 3 units of insulin on the same healthy human subject subsisting on (A) a low-carbohydrate high-fat diet and (B) a high-carbohydrate low-fat diet. Note the greater sensitivity to insulin in the latter case and the shortened latent period before the blood sugar begins to fall.

(*After* H. P. Himsworth (1935–38) *Clin. Sci.* **2,** 81.)

In the body insulin accelerates the conversion of glucose to glycogen in the muscles. It is not essential for this process, since glycogen deposition can occur in the muscles even of pancreatectomized animals, and since the synthesis of glycogen *in vitro* by isolated enzyme systems is not accelerated by insulin. Nevertheless both in the normal and in the depancreatized animal increased glycogen deposition can be demonstrated as the result of insulin administration. If sufficient glucose is supplied to normal animals receiving insulin by injection, still greater amounts of glycogen are deposited in the muscles. The action of insulin on muscle glycogenesis may also be demonstrated *in vitro.* The extent to which insulin influences glycogen deposition in muscle varies with the concentration of sugar. At high sugar concentrations (for example 400 mg. per 100 ml.)

insulin exerts little or no effect on the amount of glycogen deposited in isolated rat diaphragm, but at normal sugar concentrations (100 mg. per 100 ml.) more than twice as much glycogen is deposited in the presence as in the absence of insulin. In the muscles, therefore, insulin accelerates glycogenesis from the blood sugar. Glycogenesis from lactic acid (p. 298) does not appear to be affected by insulin since it occurs normally in the depancreatized animal. There is no evidence that insulin has any inhibitory effect on glycolysis in muscle. In the liver, on the other hand, insulin inhibits glycogenolysis, since no fall in liver glycogen follows the simultaneous injection of both adrenaline and insulin.

Insulin is known to accelerate the oxidation of carbohydrate in muscle tissue, but it is not yet clear whether it acts directly on one of the steps in the pyruvate oxidation cycle or in a more indirect manner.

As should already be clear, the key position previously attributed to the pancreas in the metabolism of carbohydrate is being challenged by much recent work. For example, if the pancreas is removed from a dog, and if insulin is continuously injected at a rate just sufficient to maintain blood sugar at the normal level, an experimental preparation is obtained which differs from the normal dog in its inability to produce additional insulin as required. When the glucose tolerance test is carried out on such an animal a perfectly normal curve is obtained, showing that no additional secretion of insulin is required in order to obtain a normal glucose tolerance curve provided sufficient insulin is present to maintain a normal blood-sugar level. On the other hand, if the liver is removed from a healthy dog, and if the blood sugar is kept at the normal level by continuous intravenous infusion of glucose, a preparation is obtained from which, after a test dose of glucose, a typical diabetic blood-sugar curve is obtained. Since such an animal is presumably able to mobilize insulin as required, it appears that the liver plays an indispensable part in the regulation of the blood-sugar level. It is significant also that patients from whom the whole pancreas has been removed for malignant disease require no more than 30 to 40 units of insulin a day, far less than is needed by many diabetic subjects.

(*b*) **Glucagon, the Hyperglycæmic-Glycogenolytic Factor.**—The α-cells of the pancreatic islets produce a hyperglycæmic-glycogenolytic factor (H.G. factor) which can raise the blood-sugar level. It is a protein and a contaminant of most insulin preparations but its effect is normally overshadowed by that of the accompanying insulin which opposes it. Its action, like that of adrenaline, is to increase the concentration of active phosphorylase in the cell.

Whether the glucagon normally plays a part in the control of the blood-sugar level is not fully understood but recent work by F. G. Young has demonstrated that injection of pituitary growth hormone into alloxan diabetic-hypophysectomized-adrenalectomized rats is followed by the appearance of glucagon in the portal blood.

3. The **adrenal medulla** produces *adrenaline* and *noradrenaline* which raise the blood sugar level by stimulating glycogenolysis. Hence the blood sugar is raised in those conditions which lead to the release of adrenaline by the suprarenal glands, for example pain, fear, anger, or asphyxia. This rise in blood sugar ensures an ample supply of carbohydrate being carried by the blood to the muscles for the muscular activity of fight or flight which follows the emotional stimulus. It is probable

that the adrenaline mechanism represents an emergency system which plays but little part in normal circumstances.

Adrenaline stimulates glycogenolysis in both liver and muscle. As the result of liver glycogenolysis, the injection of adrenaline is followed not only by hyperglycæmia but by a fall in liver glycogen. The blood lactic acid also rises as the result of increased breakdown of muscle glycogen. As we have seen, the part played by adrenaline in glycogenolysis is to accelerate the action of phosphorylase *in vivo* (but not *in vitro*).

4. **The Hormones of the Adrenal Cortex in Carbohydrate Metabolism.**—The fasting adrenalectomized animal is unable to retain any glycogen in its liver, but glycogen may be deposited if food is ingested and the proper balance of inorganic ions is maintained. If, however, crude extract of the adrenal cortex is administered to the adrenalectomized animal glycogen is deposited in the liver, even during fasting. Of the many steroids (Chapter 51) in such crude extracts only those which have an oxygen atom at position 11 in the steroid ring system are active in this respect. Thus corticosterone is active, while deoxycorticosterone is not.

5. **Other Endocrine Influences.**—Although marked disturbances of carbohydrate metabolism are found in thyroidectomized animals, and in patients with hyperthyroidism and myxœdema, there is at present no evidence that the thyroid hormone directly influences the metabolism of carbohydrate. It is more probable that the changes are to be attributed to the stimulation or retardation of metabolism which follows excess or deficiency of the hormone (Chapter 51).

GLYCOSURIA

When urine gives positive tests for reducing sugar with Benedict's or Fehling's reagent it is highly probable that the sugar is glucose and the condition *glycosuria*, although in certain circumstances reducing sugars other than glucose may be present. Lactose, for example, is frequently found in the urine of pregnant or lactating women and galactosuria has been reported in breast-fed infants. Galactose and fructose (lævulose) may appear in the urine during the liver function tests in which they are used, and either sugar may be found persistently in the urine of persons with rare congenital anomalies of carbohydrate metabolism. Finally pentoses may appear in the urine after ingestion of large amounts of certain fruits, and very occasionally as an inborn error of metabolism peculiar to the Jewish race.

Although the concentration of sugar in the blood of healthy people varies between about 80 and 150 mg./100 ml. only very small amounts of reducing substance are found in the urine so that the usual clinical tests for glycosuria are negative. The explanation for this is that though sugar is freely filtered from the blood as it passes through the glomeruli in the kidney (Chapter 34) it is subsequently almost completely reabsorbed by the renal tubules. The capacity of the tubules to reabsorb glucose is, however, limited, and when the concentration of glucose in the blood rises above about 160 mg. per 100 ml. the maximum tubular capacity is exceeded and glycosuria occurs.

The reabsorptive capacity of the renal tubules for glucose varies somewhat from one person to another, tending to be higher in persons, such as patients with diabetes mellitus, in whom the blood sugar concentration is habitually raised. In some otherwise perfectly healthy individuals (0·3 per cent. of army recruits) the reabsorptive capacity is less than normal so that glycosuria is present even when the blood sugar concentration is within normal limits (Fig. 20, 12). The maximum tubular reabsorptive capacity for glucose is often reduced temporarily in pregnancy during which glucosuria may occur for this reason.

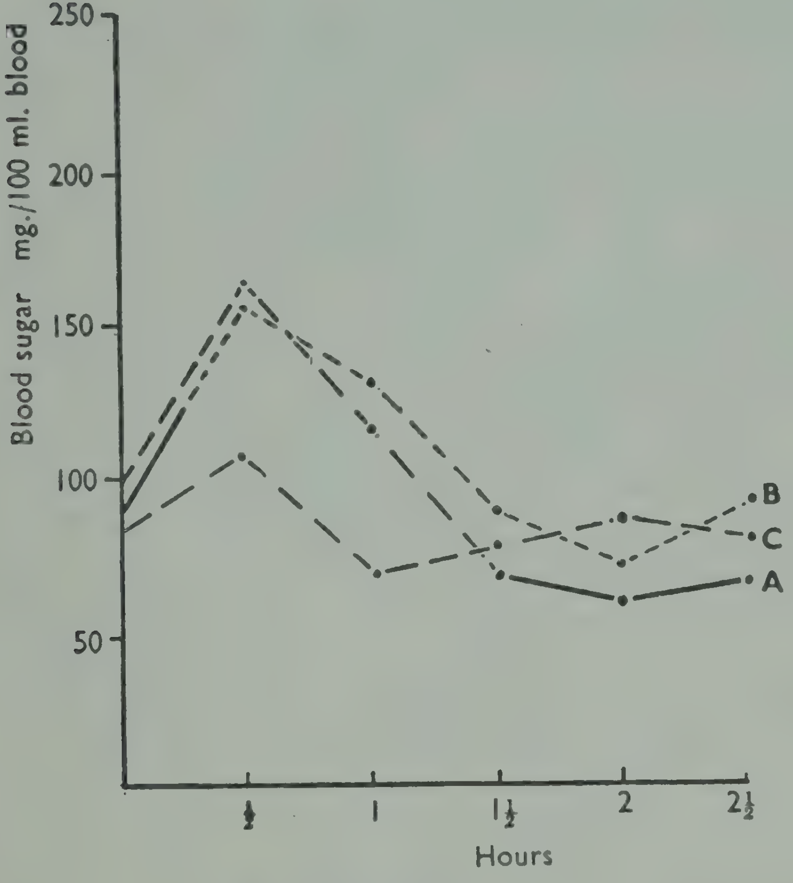


FIG. 20, 12

Blood-sugar curves (venous) after 50 g. glucose in three subjects with renal glycosuria. Interrupted line = periods of glycosuria.

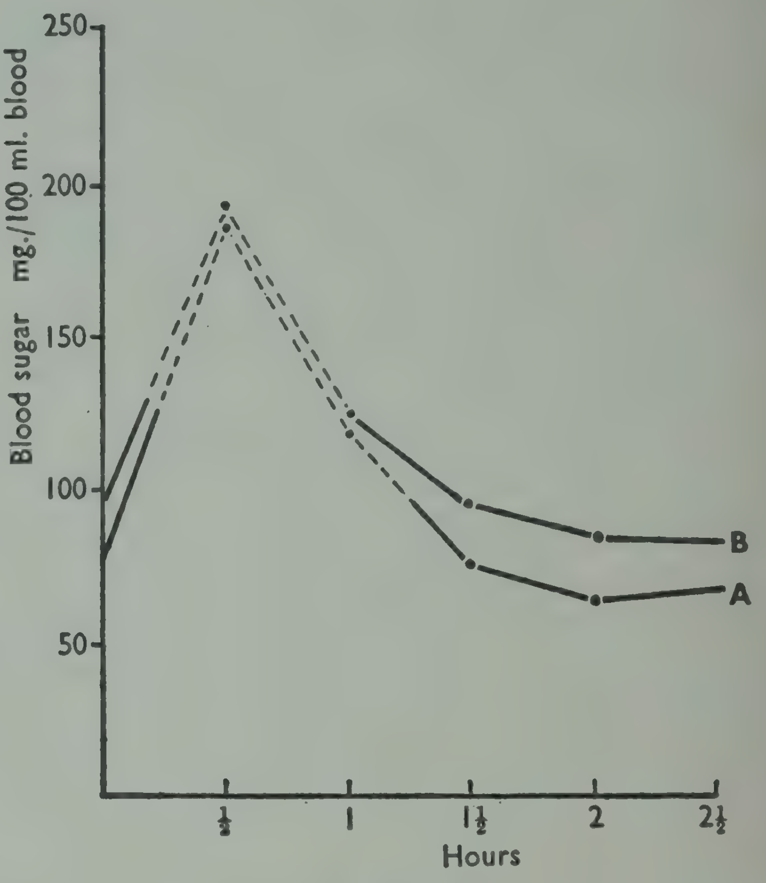


FIG. 20, 13

Blood-sugar curves (venous blood) in two subjects after 50 g. glucose. Note the unusually high peak in the ½ hour specimen and the subsequent rapid fall. Periods of glycosuria are indicated by interrupted lines.

We have already seen that the concentration of sugar in the blood is very much influenced by the rate at which sugar enters the blood from the alimentary tract and the rate at which it is removed from the blood by the cells of the liver. Sometimes, particularly after surgical operations involving removal of large parts of the stomach, glucose may be absorbed into the blood at a rate greater than that at which, initially, the liver and muscles can convert it into glycogen; a temporary hyperglycæmia and glycosuria therefore occurs (Fig. 20, 13). A similar phenomenon is occasionally observed in persons whose daily intake of carbohydrate food has been exceptionally low for some time, for example as the result of taking a ‘reducing’ diet.

Transient hyperglycæmia, and therefore glycosuria, may follow emotional excitement. It is observed, for example, in students sitting examinations, at medical examinations for insurance purposes or for National Service, and in similar conditions of mental stress. This temporary disturbance of carbohydrate metabolism is the result of stimulation of the adrenal medulla through the sympathetic nervous system and the consequent liberation of adrenaline. As we have already seen adrenaline increases the blood sugar concentration by stimulating glycogenolysis.

The French physiologist, Claude Bernard, found that puncture of the floor of the fourth ventricle in the brain of the rabbit was followed by hyperglycæmia and glucosuria, provided that the animal had adequate stores of glycogen in the liver (diabetes piqûre). Here again the hyperglycæmia is brought about by stimulation of the adrenal glands through the sympathetic nervous system, for it is not observed if the splanchnic nerves are cut or the adrenal gland removed. The glycosuria that sometimes occurs in patients with intracranial disease may be produced in a similar way.

Experimental Glycosuria.—In animals glycosuria may be produced experimentally by certain operative procedures, by injection of anterior pituitary extracts (p. 300) and by certain drugs. Such experiments have made an important contribution to our knowledge of carbohydrate metabolism.

(a) *Pancreatectomy* (extirpation of the pancreas), by removing the natural source of insulin, is followed by hyperglycæmia and glycosuria.

(b) *Alloxan*, a ureide derivative (p. 58), when given by injection to animals causes necrotic changes in the Islets of Langerhans and a condition resembling diabetes mellitus (Shaw Dunn). Dehydroascorbic acid has recently been shown to produce diabetes mellitus in a similar fashion.

(c) *Phlorrhizin* is a drug which causes glycosuria when taken by mouth or injected into animals or man. It does not produce hyperglycæmia, but it inhibits phosphorylase and apparently acts by reducing the ability of the renal tubule cells to reabsorb glucose from the glomerular filtrate. After the administration of phlorrhizin the blood sugar level falls as the result of the continuous excretion of glucose and the carbohydrate stores eventually become exhausted. A fasting phlorrhizinized animal attempts to maintain its blood sugar level by converting the antiketogenic amino-acids (p. 345) of its own tissue proteins into glucose (gluconeogenesis) with consequent serious loss of tissue. An index to the severity of the condition can be obtained by meauring the G/N or D/N ratio, that is the ratio of glucose (or dextrose) to nitrogen excreted in the urine. While carbohydrate stores are still available the amount of glucose relative to nitrogen is high; but as carbohydrate stores diminish, and as tissue proteins are utilized, the amount of glucose tends to fall, while the amount of nitrogen rises. The ratio therefore falls until it reaches the value of 3·65, indicating that tissue protein only is being utilized.

When protein is fed to such an animal antiketogenic amino-acids are converted to glucose which is at once excreted in the urine. Thus information can be obtained about the gluconeogenetic properties of any particular protein or amino-acid (p. 345) by studying its effect upon the D/N ratio in a phlorrhizinized animal. An amino-acid which causes an increase in the D/N ratio is obviously antiketogenic.

DIABETES MELLITUS

Glycosuria occurs in many diseases of which disordered metabolism of carbohydrate is a feature. Some of these will be referred to in Chapter 51, but we must deal now with the commonest, namely *diabetes mellitus*. This is a metabolic disorder of which impaired utilization of carbohydrate is the most obvious feature. It must be emphasized at once, however, that the metabolism of fat and of protein, and indeed the total food metabolism, is affected. The disease is worth considering in some detail because its study has yielded valuable information about normal carbohydrate metabolism.

Although diabetes has been known for many centuries it is only quite recently that detailed information as to its nature has been obtained. The work of von Mering and Minkowski in 1889 first drew attention to the relationship of the pancreas to carbohydrate metabolism. They found that dogs from which the pancreas had been removed developed a condition similar to human diabetes mellitus in which large amounts of urine containing sugar were excreted. If a piece of pancreas was grafted under the skin the condition either did not occur at all, or developed only in a much milder form, suggesting that the pancreas produced some substance which was liberated, not into the pancreatic duct, but directly into the blood. This factor is named *insulin*, because it is produced by the cells of the islet tissue in the pancreas described by Langerhans in 1869. Injections of insulin can remove the signs and symptoms of diabetes in a depancreatized animal, which may be kept in good health for long periods provided it is given a suitable diet. Injections of insulin can also correct diabetes mellitus in man by taking the place of insulin from the patient's own islet cells, but modern research has made it clear that not all cases of diabetes mellitus are primarily due to disease of the islets of Langerhans. Insulin itself is discussed in more detail in Chapter 51.

Although the exact nature of the defect in human diabetes is not yet fully understood, many of the features of the disorder can be explained on the basis of a defective insulin supply (relative or absolute). In diabetes the response to a high blood-sugar level in terms of insulin production is poor. Thus in the glucose tolerance test the diabetic gives a response which differs markedly from that of the normal subject (Fig. 20, 14). In mild diabetes the blood-sugar curve may start near the normal fasting blood-sugar level, but it rises to a higher level than does the normal curve, the peak is reached later and the fall takes place more slowly (Fig. 20, 14). During part of this time the maximum capacity of the renal tubules to reabsorb glucose is exceeded and glycosuria occurs. In severe diabetes the fasting level is usually above the renal threshold value, and glycosuria occurs continuously.

The hyperglycæmia of diabetes is associated with increased liver glycogenolysis and deficient muscle glycogenesis. Although the tissues, especially the muscles, receive a liberal supply of glucose from the blood they are unable to utilize it efficiently. As a result, the diabetic person experiences muscular weakness. Moreover he loses a large amount of potential fuel in the urine. In severe diabetes the amount of glucose lost into the urine may be very considerable (over 100 g. a day), and much water must also be excreted to keep all this glucose in solution. The diabetic patient therefore passes abnormally large amounts of urine (polyuria); he is thirsty and drinks much fluid (polydipsia). Because his tissues cannot utilize glucose, although they need fuel, he is hungry also, and frequently has a craving for sweet foods.

As carbohydrate cannot be used for fuel, some other source of energy must be found. Fat is therefore utilized. The fat stores of the body are mobilized, the fat content of the blood rises, and a sample of blood from

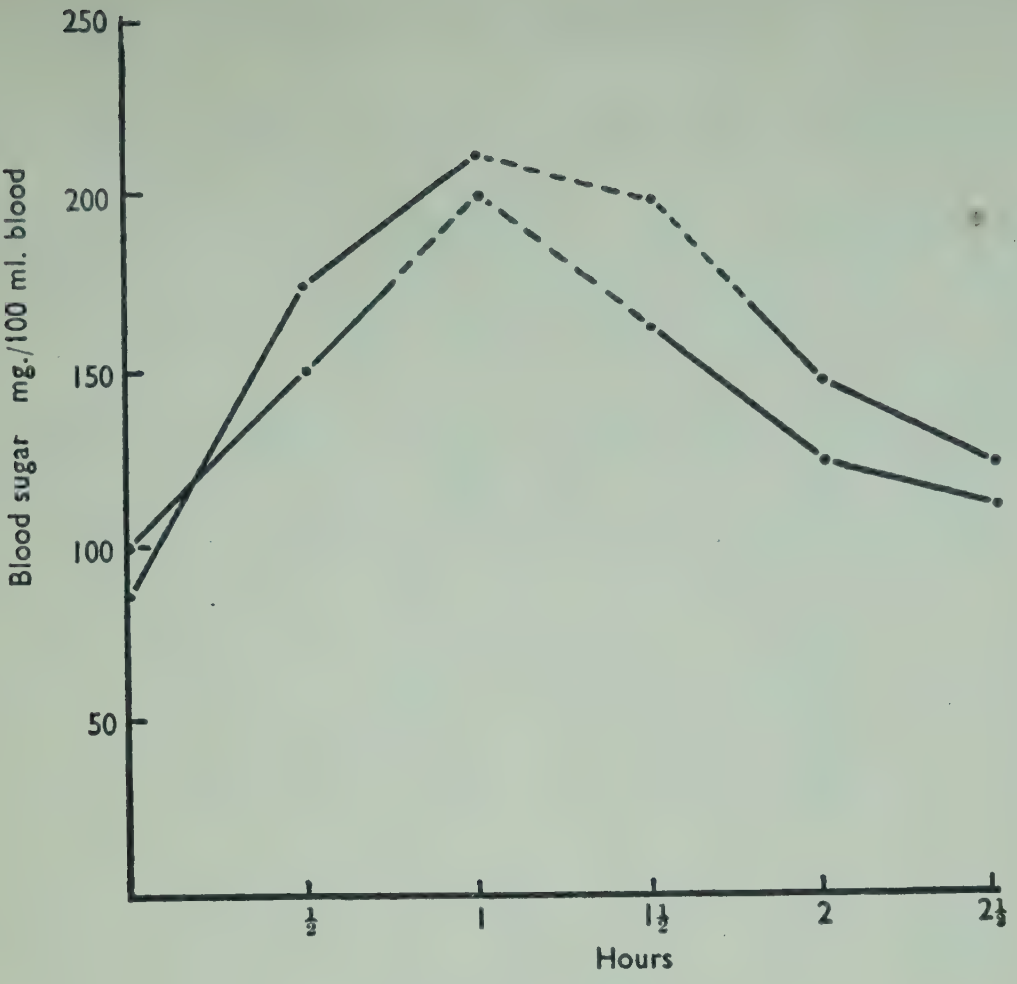


FIG. 20, 14

Blood-sugar curves from two mildly diabetic patients. The dotted lines indicate glycosuria. Although the fasting blood sugar is nearly normal the response to administration of glucose is abnormal. Compare with the normal curves in Fig. 20, 1.

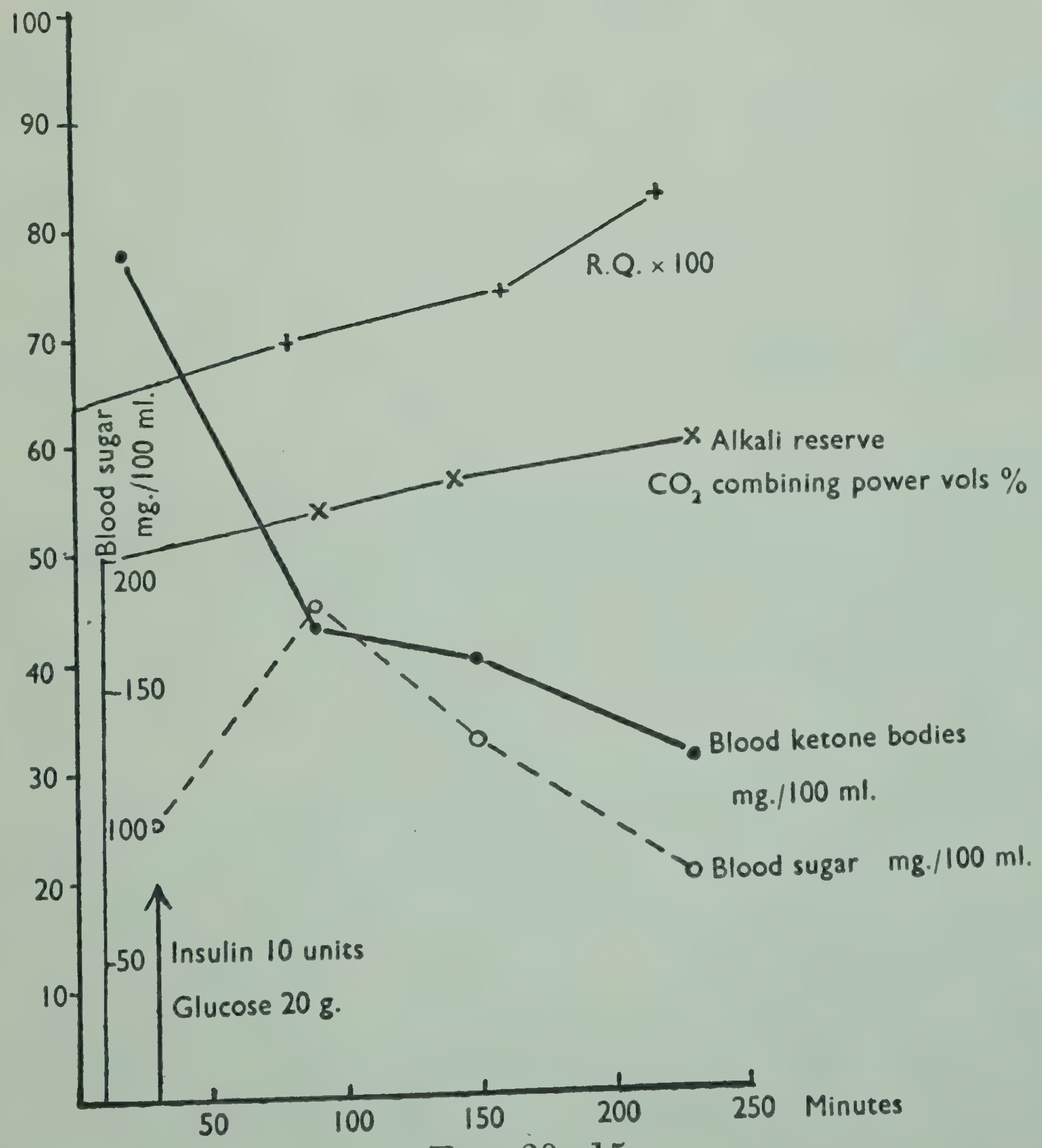


FIG. 20, 15

Changes in respiratory quotient, alkali reserve, ketonæmia and hyperglycæmia in a fasting diabetic patient following the injection of 10 units of insulin and 20 g. of glucose.

(By courtesy of D. M. Lyon.)

a case of severe diabetes often shows a thick creamy layer of fatty material on standing. Protein is also used as a source of energy with the result that in long-standing cases tissue-wasting occurs and the patient becomes weak and thin. The respiratory quotient (p. 154) of protein is 0·8 and that of fat is 0·7; the respiratory quotient, therefore, in severe diabetes is between 0·7 and 0·8.

In a vain attempt to build up a reserve of glycogen the antiketogenic amino-acids of both the dietary proteins and the tissue proteins are converted to glucose which, however, cannot be utilized and is lost in the urine.

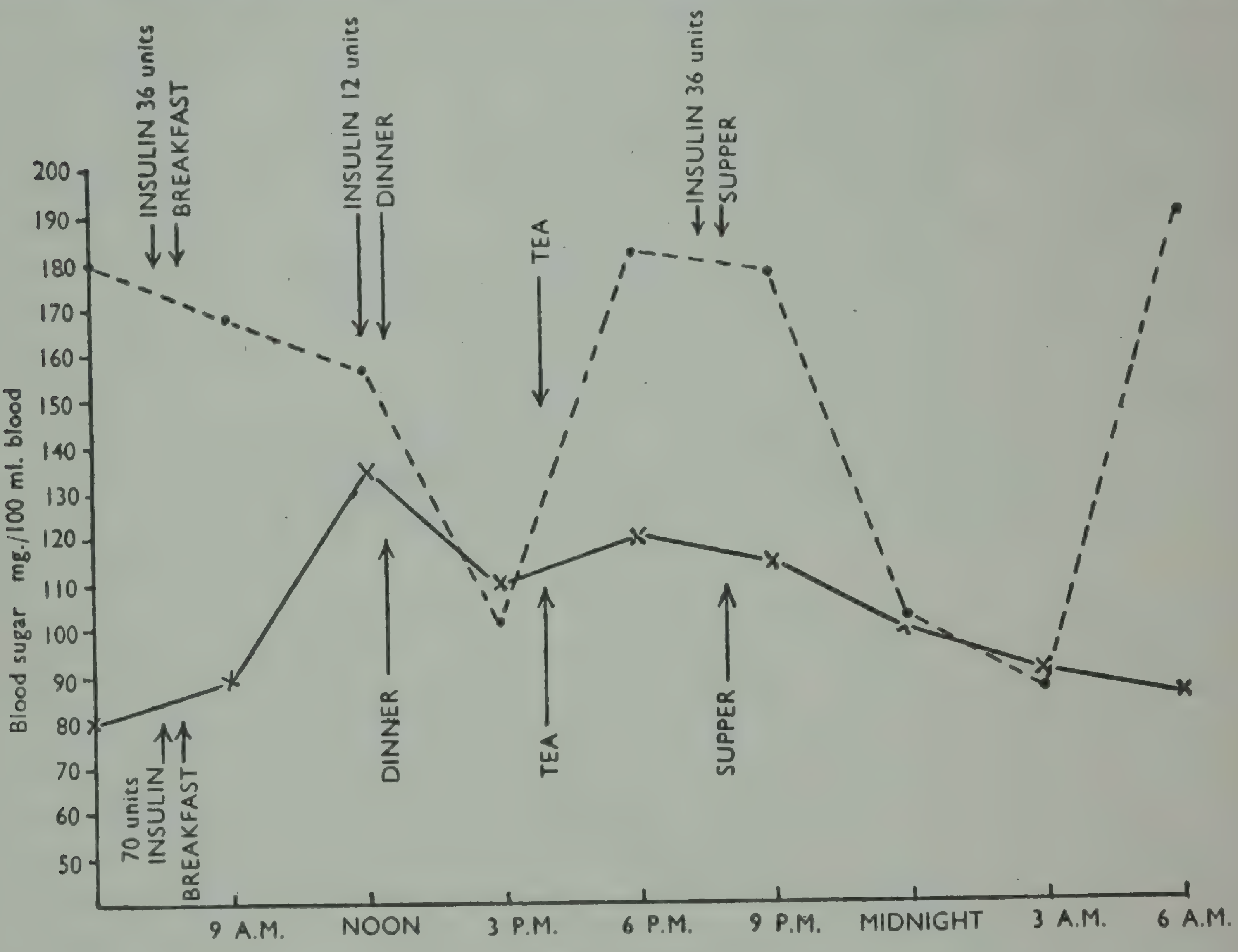


FIG. 20, 16

Capillary blood-sugar levels in a diabetic patient on a standard diet. Interrupted line, patient having three injections of soluble insulin (total 84 units). Note the oscillations in blood-sugar level, the rise during the night and the high blood sugar in the morning. Continuous line, patient having a single injection of zinc protamine insulin (70 units at 7.30 a.m.). Note the more stable blood sugar, the highest value in the morning before the full effect of the delayed action insulin has been obtained, and the gradual fall in blood sugar during the night.

The disproportionate fat metabolism in diabetes results in excessive production of ketone bodies. Ketonæmia and ketonuria (p. 332) are therefore features of severe diabetes. If the condition is severe and is not treated the patient passes into a state of diabetic coma from which he may not recover. In diabetic coma the odour of acetone may be noticeable in the patient's breath. As two of the ketone bodies are acids, the alkali reserve of the blood falls and the carbon dioxide-combining power is lowered. The respiration becomes slow and deep ('air-hunger') and the alveolar CO_2 tension falls. The titratable acidity of the urine increases as does its content of ammonium salts.

The continued loss of water from the body into the urine that occurs during the development of diabetic ketosis results in a state of 'dehydration' together with a profound disturbance of tissue electrolytes which is further dealt with in Chapter 35.

Although mild cases of diabetes may respond to dietary treatment alone, all severe cases require insulin. In the treatment of diabetes, then, insulin is injected to supply that part of the body's requirement which cannot be met by the pancreas. Since insulin is hydrolysed by proteolytic enzymes it is useless if given by mouth. The original 'soluble' insulin has to be injected several times a day, about half an hour before each meal, but preparations of insulin in combination with protamine or globin are now available from which the insulin is slowly liberated during some 16 to 20 hours after injection. These preparations exercise a more prolonged effect upon the blood sugar, so that a single injection given in the morning may suffice for twenty-four hours. A further advantage is that after their administration the blood sugar fluctuates less widely throughout the day (Fig. 20, 16).

Hypoglycæmia.—An overdose of insulin produces a fall in the blood-sugar level (hypoglycæmia), convulsions and coma which, however, is rarely fatal. The early symptoms of hypoglycæmia resemble those produced by adrenaline and have been dealt with on p. 302. From what has already been said it will be clear that hypoglycæmia can result from other causes, for example an excessive production of insulin ('hyperinsulinism'), defective glycogen stores (hepatitis and glycogen disease) and disorders of the pituitary gland.

GLYCOLYSIS

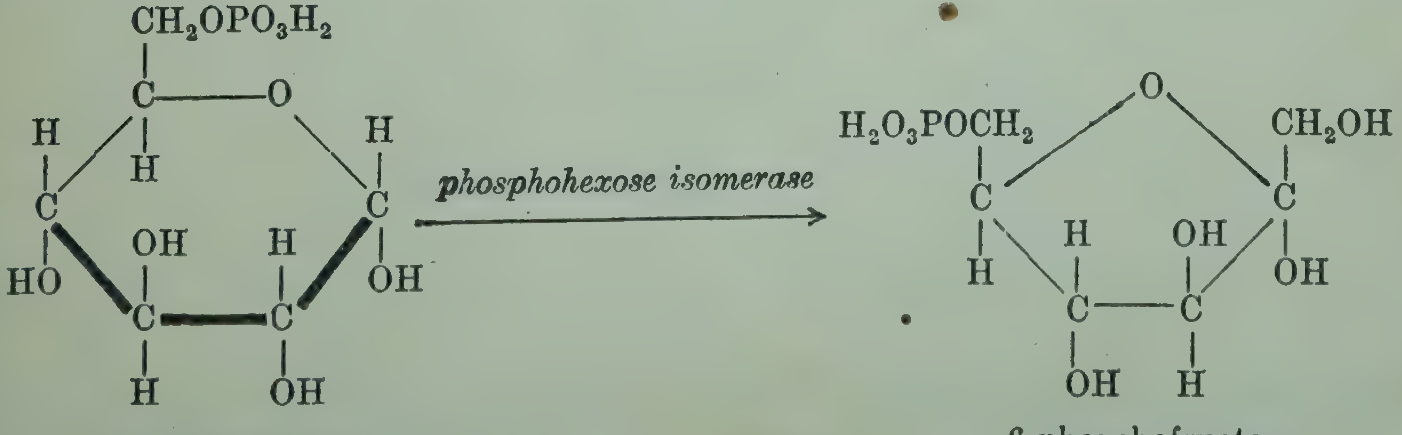
THE ANAEROBIC BREAKDOWN OF CARBOHYDRATE

It has already been mentioned that in certain tissues, particularly in muscle, glycogen breaks down to yield pyruvic acid in the process known as *glycolysis*. Although this term was originally restricted to glycogen breakdown in muscle it is now fairly generally applied to the anaerobic breakdown of carbohydrate (glycogen or glucose) with the production of pyruvate and the liberation of energy.

The process of glycolysis is elaborate and takes place in many stages, a simplified account of which is presented in the following pages. The initial steps whereby glycogen is converted to 6-phosphoglucose have already been discussed (Fig. 20, 7).

$$\text{glycogen} \overset{\textit{phosphorylase}}{\rightleftharpoons} \text{1-phosphoglucose} \overset{\textit{phosphoglucomutase}}{\rightleftharpoons} \text{6-phosphoglucose}$$

Under the influence of the enzyme *oxoisomerase* (or *phosphohexose isomerase*) some of the 6-phosphoglucose becomes converted into the corresponding derivative of fructose,

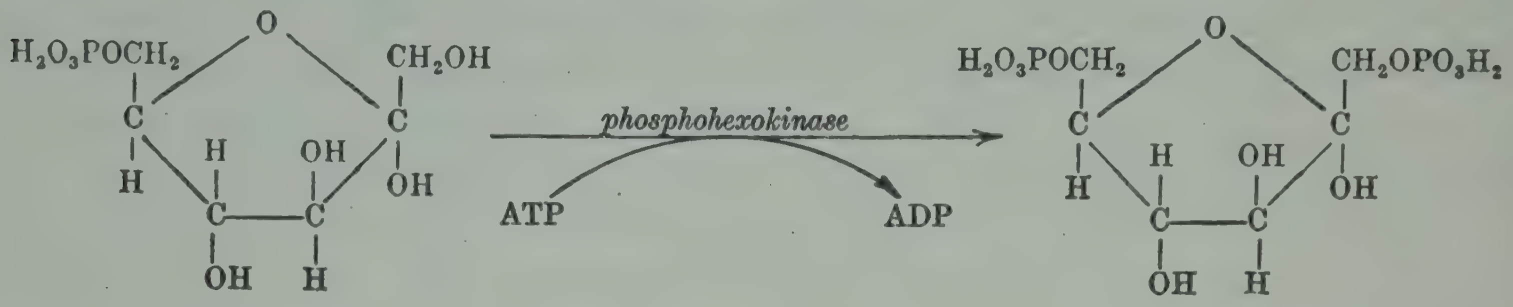


6-phosphoglucose

6-phosphofructose

6-phosphofructose (the Neuberg ester) in which fructose is present in the furanose form (p. 10). An equilibrium mixture of the esters is formed, consisting of 70 per cent. 6-phosphoglucose and 30 per cent. 6-phosphofructose. In subsequent stages the latter is used up, so that eventually all the 6-phosphoglucose becomes 6-phosphofructose.

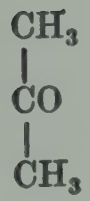
In the next stage a second molecule of phosphoric acid is introduced into the fructose ring, this time in position 1, so that the compound now formed is *1:6 diphosphofructose* (the Harden and Young ester). This reaction occurs under the influence of the enzyme *phosphohexokinase* and the source of the necessary phosphate is adenosine triphosphate (ATP) which thereby becomes converted to adenosine diphosphate (ADP).



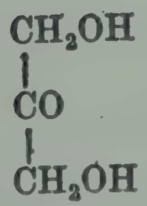
6-phosphofructose

1, 6-diphosphofructose

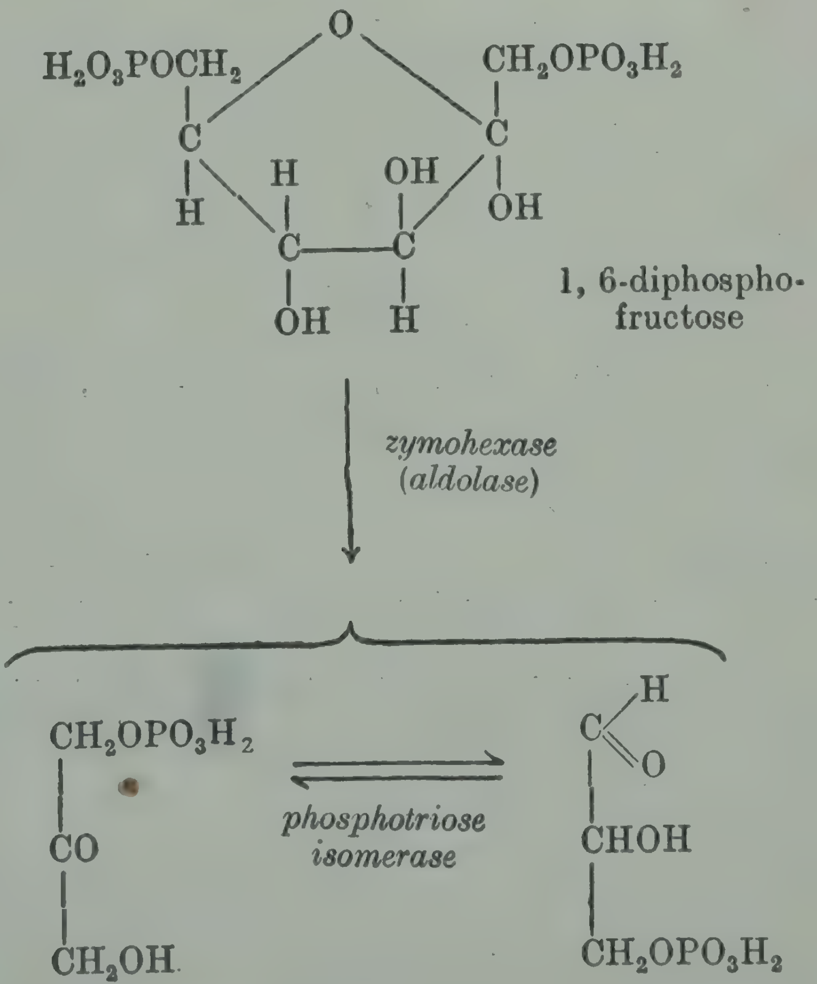
The diphosphofructose molecule now splits under the influence of the enzyme *zymohexase* (*aldolase*) into two smaller molecules each containing three carbon atoms and one phosphorus atom. One of these products is *1-phosphodihydroxyacetone*; the other is *3-phosphoglyceraldehyde*. The former is a phosphoric ester of the ketose, dihydroxyacetone, while the latter is a phosphoric ester of the aldose sugar, glyceraldehyde. These sugars are both trioses and the two phosphoric esters are therefore frequently termed phosphotrioses or triosephosphates. They are formed from diphosphofructose in equimolecular amounts, but under the influence of the enzyme *isomerase* (*phosphotriose isomerase*) the one can be converted into the other and an equilibrium is reached in which *phosphodihydroxyacetone* predominates. Subsequently 3-phosphoglyceraldehyde is used up as described later and the reaction therefore proceeds from left to right in the diagram immediately below.


$$\begin{array}{c} CH_3 \\ | \\ CO \\ | \\ CH_3 \end{array}$$

acetone

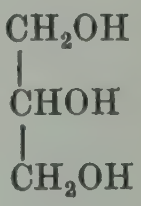

$$\begin{array}{c} CH_2OH \\ | \\ CO \\ | \\ CH_2OH \end{array}$$

dihydroxyacetone

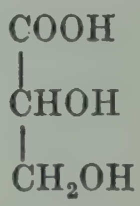


1-phosphodihydroxy-acetone

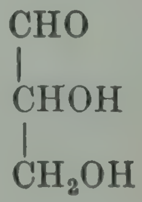
3-phospho-glyceraldehyde


$$\begin{array}{c} CH_2OH \\ | \\ CHOH \\ | \\ CH_2OH \end{array}$$

glycerol

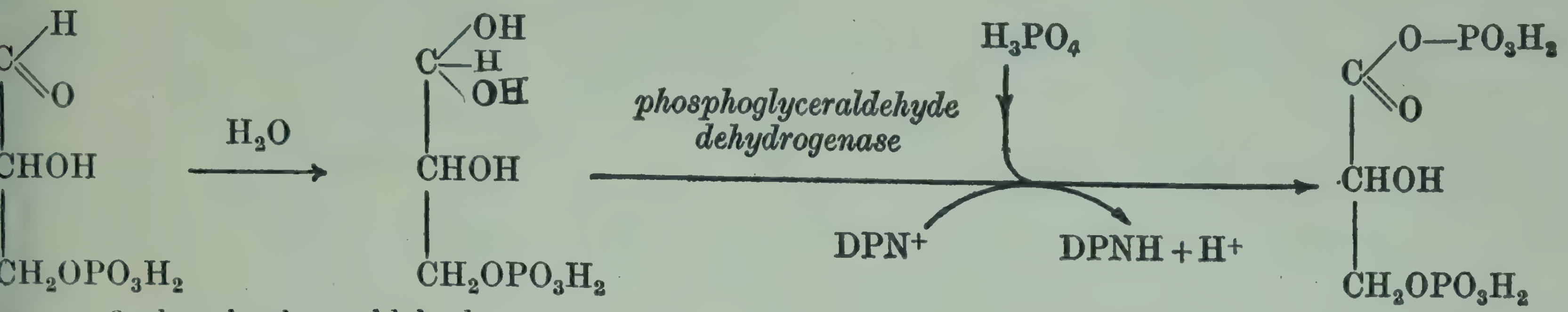

$$\begin{array}{c} COOH \\ | \\ CHOH \\ | \\ CH_2OH \end{array}$$

glyceric acid


$$\begin{array}{c} CHO \\ | \\ CHOH \\ | \\ CH_2OH \end{array}$$

glyceraldehyde

In the next stage phosphoglyceraldehyde in its hydrated form is oxidized to *1 : 3-diphosphoglyceric acid* under the influence of the enzyme *phosphoglyceraldehyde dehydrogenase* (*triosephosphate dehydrogenase*) which has glutathione as its prosthetic group. This involves the removal of hydrogen which is transferred to cozymase (diphosphopyridine nucleotide, DPN). At the same time a new phosphate group is introduced in position 1. This group is 'energy-rich.'

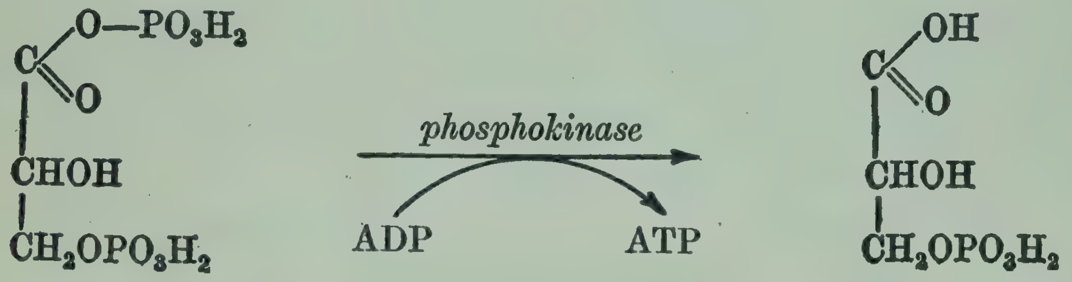


3-phosphoglyceraldehyde

1 : 3-diphospho-glyceric acid

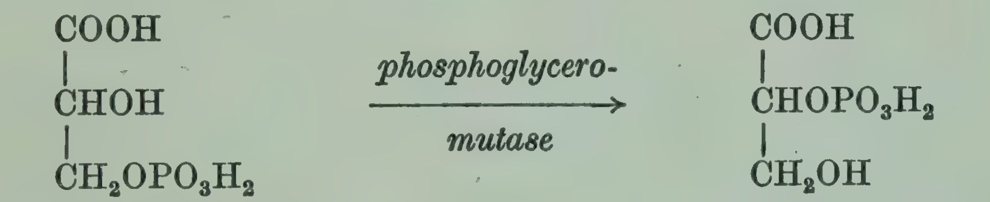
This reaction is inhibited by iodoacetic acid because iodoacetic acid combines with the sulphydryl groups of the glutathione, the prosthetic group of the enzyme.

The 'energy rich' phosphate group in position 1 is now transferred to adenosine-diphosphate (ADP) thus:



This reaction is catalysed by a *phosphokinase* and results in the formation of ATP.

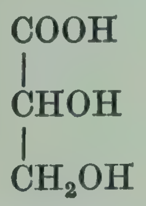
The phosphate grouping now moves from position 3 to position 2, so that *3-phosphoglyceric acid* becomes converted to *2-phosphoglyceric acid.* This reaction occurs under the influence of the enzyme *phosphoglyceromutase*:



3-phosphoglyceric acid

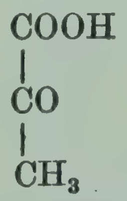
2-phosphoglyceric acid

The next step involves the conversion of this derivative of glyceric acid to a derivative of the enol form of pyruvic acid:



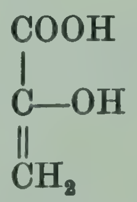
$$\begin{array}{l}\mathrm{COOH}\\ |\\ \mathrm{CHOH}\\ |\\ \mathrm{CH_2OH}\end{array}$$

glyceric acid



$$\begin{array}{l}\mathrm{COOH}\\ |\\ \mathrm{CO}\\ |\\ \mathrm{CH_3}\end{array}$$

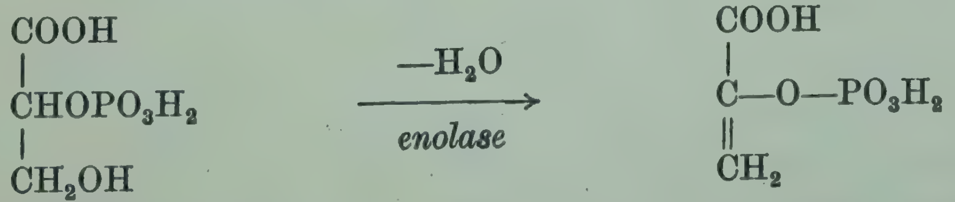
pyruvic acid



$$\begin{array}{l}\mathrm{COOH}\\ |\\ \mathrm{C{-}OH}\\ \|\\ \mathrm{CH_2}\end{array}$$

enol form of pyruvic acid

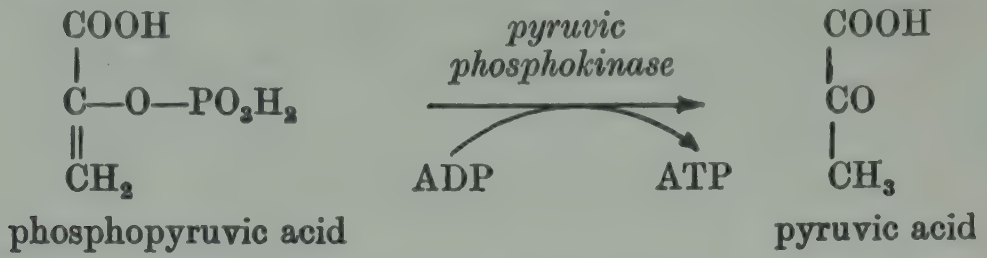
Under the influence of the enzyme *enolase*, 2-phosphoglyceric acid is converted into *(enol)-phosphopyruvic acid* and the phosphate bond becomes 'energy-rich.' This reaction is inhibited by fluorides:



2-phosphoglyceric acid

(enol)-phosphopyruvic acid

At this stage the long association of carbohydrate derivatives with phosphoric acid comes to an end. Phosphopyruvic acid gives up its phosphate grouping which is taken up by adenosinediphosphate (ADP) to form adenosine triphosphate (ATP). Free pyruvic acid is produced and the store of ATP is built up again.

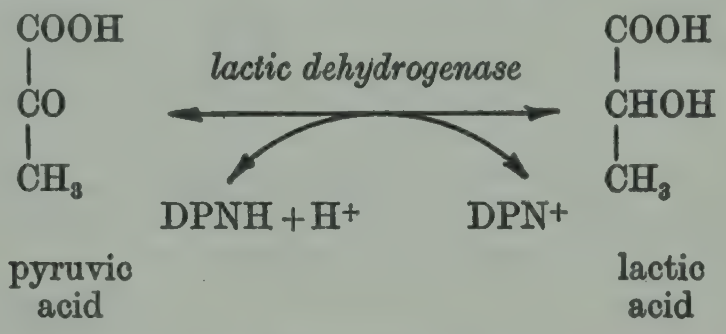


$$\mathrm{CH_2{=}C(O{-}PO_3H_2){-}COOH} \xrightarrow[\mathrm{ADP}\,\rightarrow\,\mathrm{ATP}]{\textit{pyruvic phosphokinase}} \mathrm{CH_3{-}CO{-}COOH}$$

The fate of the pyruvate which has been formed in this elaborate process of glycolysis depends upon the amount of oxygen available.

If abundant oxygen is present the pyruvate is oxidized to carbon dioxide and water in the citric acid cycle which is discussed in the next section. At the same time the DPN, which was reduced by taking up a pair of hydrogen atoms when the phosphoglyceraldehyde was oxidized to phosphoglyceric acid, is oxidized again by the flavoprotein and cytochrome systems and ultimately, of course, by molecular oxygen (Chapter 8).

If the oxygen supply is deficient, oxidation of reduced DPN in this way is not possible, nor is it possible for pyruvic acid to be oxidized by the citric acid cycle. In these circumstances pyruvic acid is reduced to lactic acid at the expense of the reduced DPN thus:



$$\mathrm{CH_3{-}CO{-}COOH} \xrightleftharpoons[\mathrm{DPNH+H^+}\,\rightarrow\,\mathrm{DPN^+}]{\textit{lactic dehydrogenase}} \mathrm{CH_3{-}CHOH{-}COOH}$$

This reaction, which is catalysed by the enzyme *lactic dehydrogenase*, is reversible and is in fact reversed when lactic acid is ultimately oxidized in muscle during recovery after exercise.

The complete scheme of glycolysis is summarized in Fig. 20, 17, which emphasizes the importance of ATP in the process. In the course of the conversion of each 6-carbon unit of glycogen to pyruvic acid (or lactic acid) one molecule of ATP is used up in the phosphorylation of 6-phosphofructose while four molecules of ATP are produced (two in the dephosphorylation of diphosphoglyceric acid and two in the dephosphorylation of phosphopyruvic acid). There is thus a net gain of three molecules of ATP and the process of glycolysis may be regarded essentially as a mechanism for the anaerobic production of ATP.

A series of chemical changes very similar to those involved in the breakdown of glycogen to pyruvic acid in muscle takes place when carbohydrate is converted to *alcohol* in the process of *alcoholic fermentation* by yeast. In the latter case the starting material may either be starch, which is broken down to glucose, or glucose. In both cases the reaction proceeds as in Fig. 20, 17, by way of 6-phosphoglucose, the triose phosphates, etc., until the stage of pyruvic acid is reached. In yeast fermentation the pyruvic acid is decarboxylated, by removal of carbon dioxide under the influence of the enzyme *carboxylase* and its coenzyme diphosphothiamine, to yield acetaldehyde which accepts two hydrogen atoms to yield alcohol.

In alcoholic fermentation, therefore, ethyl alcohol and carbon dioxide are the end-products of carbohydrate breakdown.

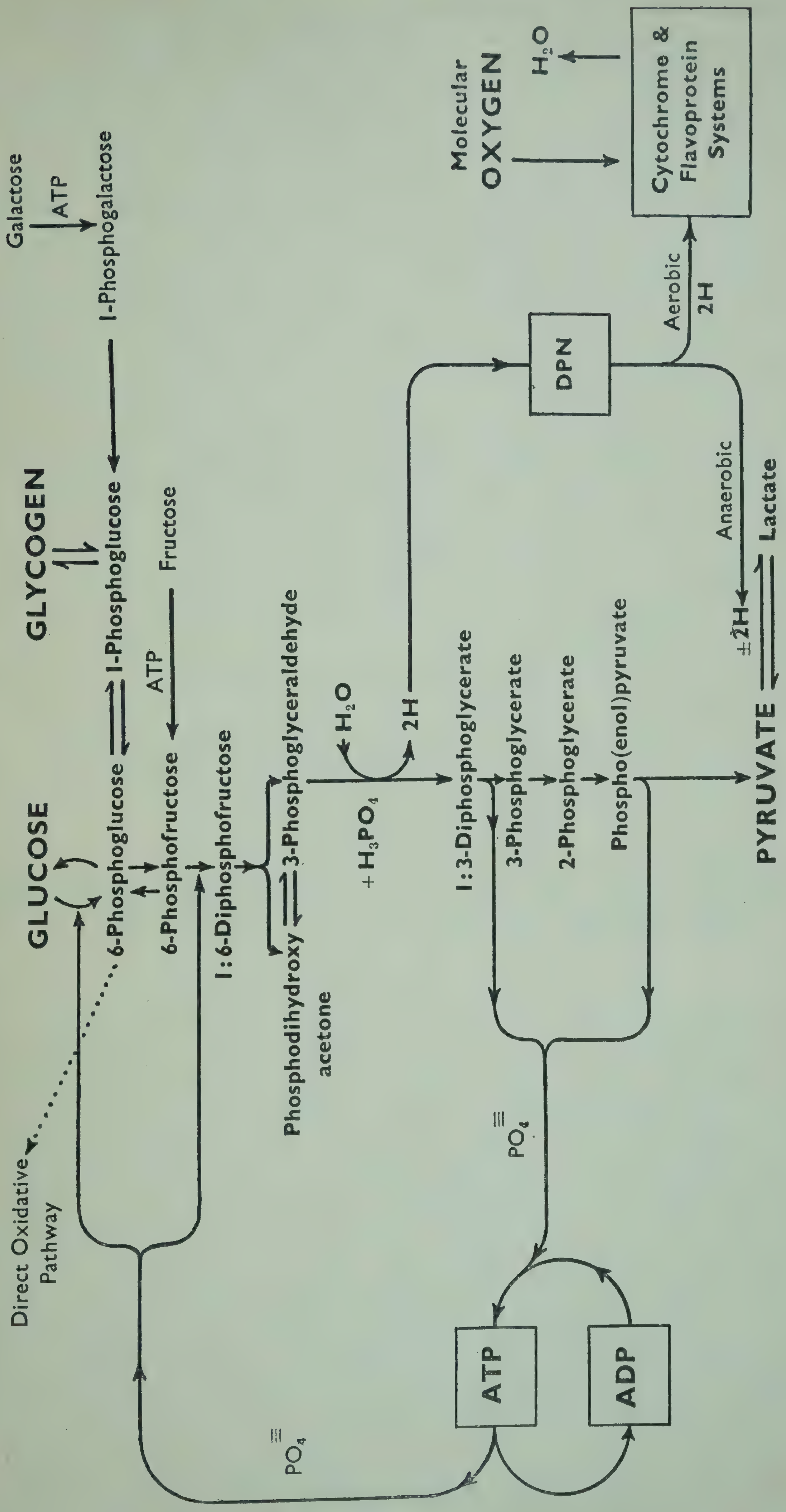

Galactose
ATP
1-Phosphogalactose
GLYCOGEN
1-Phosphoglucose
Fructose
ATP
GLUCOSE
6-Phosphoglucose
6-Phosphofructose
1:6-Diphosphofructose
Phosphodihydroxy acetone
3-Phosphoglyceraldehyde
H_2O
$+ H_3PO_4$
2H
1:3-Diphosphoglycerate
3-Phosphoglycerate
2-Phosphoglycerate
Phospho(enol)pyruvate
PYRUVATE
±2H
Lactate
Anaerobic
Aerobic
2H
DPN
Molecular OXYGEN
H_2O
Cytochrome & Flavoprotein Systems
Direct Oxidative Pathway
$PO_4^{\equiv}$
$PO_4^{\equiv}$
ATP
ADP

FIG. 20, 17

The Pasteur Effect.—Alcoholic fermentation and muscle glycolysis are obviously two essentially similar processes for the liberation of energy by the breakdown of carbohydrate. The living cell may obtain energy from carbohydrate molecules by two different methods, either by burning them (*oxidation*) or by fragmenting them (*fermentation* or *glycolysis*). If the cell oxidizes the carbohydrate molecule to water and carbon dioxide, it obtains all the energy stored in the carbohydrate. But if it splits the carbohydrate molecule into lactic acid (or pyruvic acid, or alcohol and carbon dioxide), then the residual fragments are still potential sources of energy and the cell has not made full use of all the energy present in the original carbohydrate. Fermentation or glycolysis is therefore uneconomical, since it yields much less energy from each molecule of starting material than does oxidation. It is a more primitive process than is oxidation and it can maintain only simple forms of life, such as certain types of bacteria.

In many types of cell the process of fermentation or glycolysis still persists side by side with oxidation, and some cells can make use of either process at will. Yeast, for example, can oxidize carbohydrate to water and carbon dioxide in the presence of oxygen, and can ferment carbohydrate to alcohol and carbon dioxide in the absence of oxygen. Pasteur was able to show that when oxygen was excluded much more sugar was broken down than when oxygen was allowed access (*the Pasteur effect*). Whether oxygen is present or absent the same amount of energy is produced by the cells, but in the former case the cells break down carbohydrate completely and less carbohydrate is necessary to provide the same amount of energy. The Pasteur effect is shown by many types of animal cell as well as by yeast and bacteria.

Although in most animal cells oxidation occurs in the presence of oxygen and glycolysis in its absence, in certain cases glycolysis occurs even in the presence of oxygen (*aerobic glycolysis*). The cells of malignant tumours are characterized by a high rate of aerobic glycolysis.

While in the normal scheme of glycolysis 6-phosphoglucose is converted into 6-phosphofructose, it is also possible for 6-phosphoglucose to be directly oxidized to 6-phosphogluconic acid. This oxidation may take place in the red blood cell for example, and is brought about by an enzyme system involving triphosphopyridine nucleotide (TPN). This substance is also required for further stages in the oxidation of the phosphogluconic acid, but the full significance of this so-called 'direct oxidative pathway' is beyond the scope of this book.

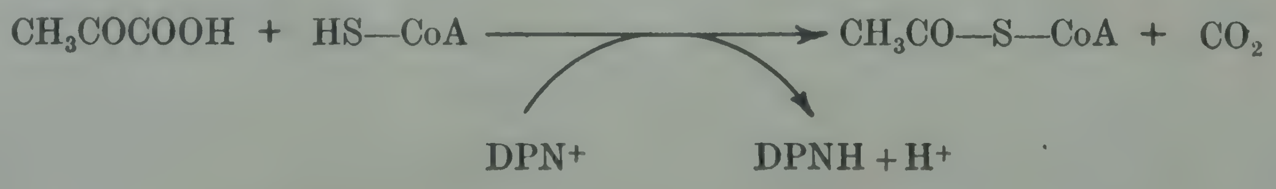
THE CITRIC ACID CYCLE (TRICARBOXYLIC ACID CYCLE)

THE AEROBIC BREAKDOWN OF CARBOHYDRATE

The Terminal Oxidation Pathway

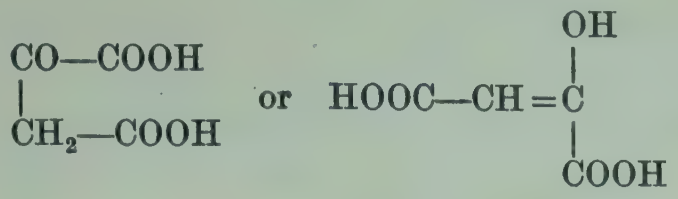
It has already been mentioned that in the presence of adequate oxygen pyruvate is oxidized to carbon dioxide and water. This process takes place by the citric acid cycle or tricarboxylic acid cycle of Krebs which operates in the mitochondria. As the details of the cycle are complex only a simplified account of the essential features will be given here.

The first step involves the participation of coenzyme A the structure of which has already been discussed on p. 62. Since the active area of its molecule is the sulphydryl($-SH$) group, the coenzyme is usually denoted by the contraction $HS-CoA$. When coenzyme A reacts with pyruvic acid, acetyl coenzyme A ($CH_3CO-S-CoA$), an energy-rich compound, is formed. At the same time cozymase (DPN^+) is reduced and carbon dioxide is liberated. The process is elaborate and probably involves the participation of diphosphothiamine and the sulphur-containing substance lipoic acid, but it may be summarized thus:



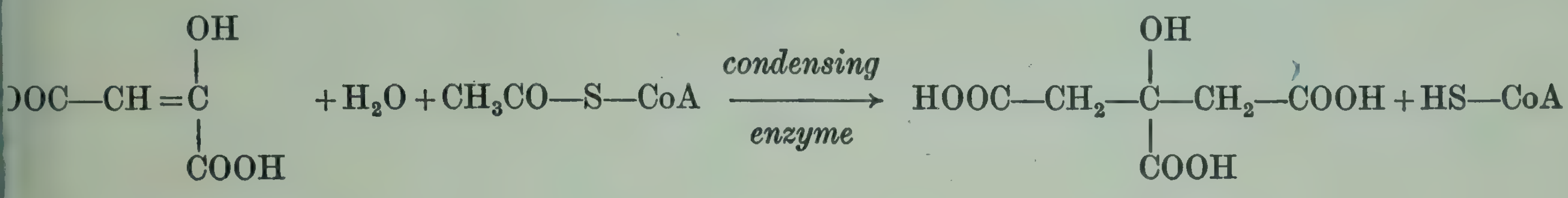
$$CH_3COCOOH + HS{-}CoA \xrightarrow[DPN^+ \;\rightarrow\; DPNH + H^+]{} CH_3CO{-}S{-}CoA + CO_2$$

Under the influence of the *condensing enzyme* the acetyl coenzyme A combines with oxaloacetic acid



$$\begin{array}{l}CO\text{—}COOH\\ |\\ CH_2\text{—}COOH\end{array} \quad \text{or} \quad HOOC\text{—}CH{=}\overset{\overset{OH}{|}}{\underset{\underset{COOH}{|}}{C}}$$

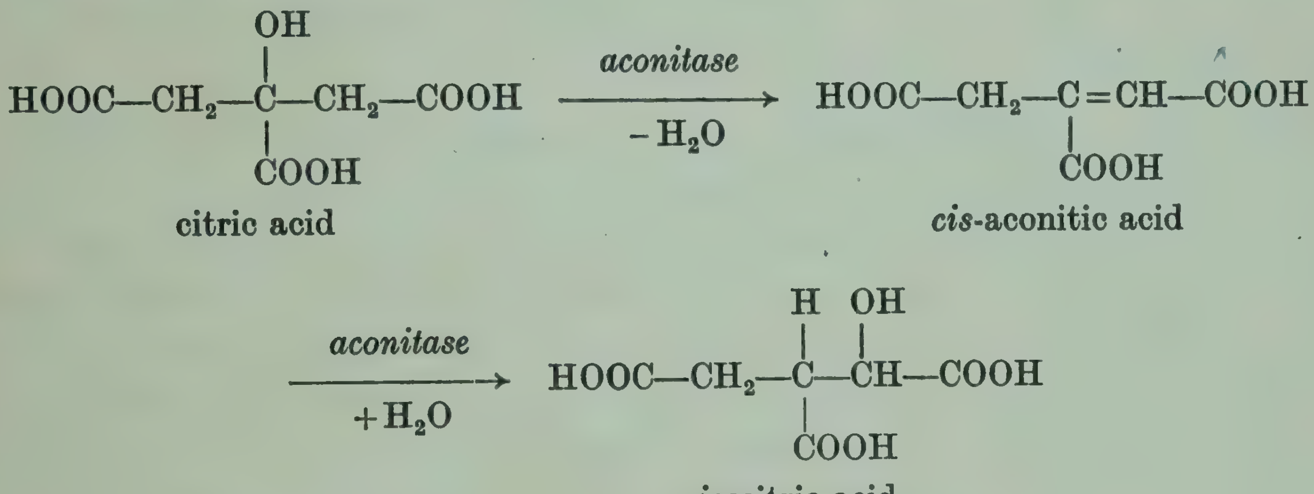
to form the tricarboxylic acid, citric acid, coenzyme A being liberated.



$$\text{OC—}CH{=}\overset{\overset{OH}{|}}{\underset{\underset{COOH}{|}}{C}} + H_2O + CH_3CO\text{—}S\text{—}CoA \xrightarrow[\textit{enzyme}]{\textit{condensing}} HOOC\text{—}CH_2\text{—}\overset{\overset{OH}{|}}{\underset{\underset{COOH}{|}}{C}}\text{—}CH_2\text{—}COOH + HS\text{—}CoA$$

oxaloacetic acid citric acid

The enzyme *aconitase* now converts citric acid to *iso*citric acid by way of the intermediate *cis*-aconitic acid.



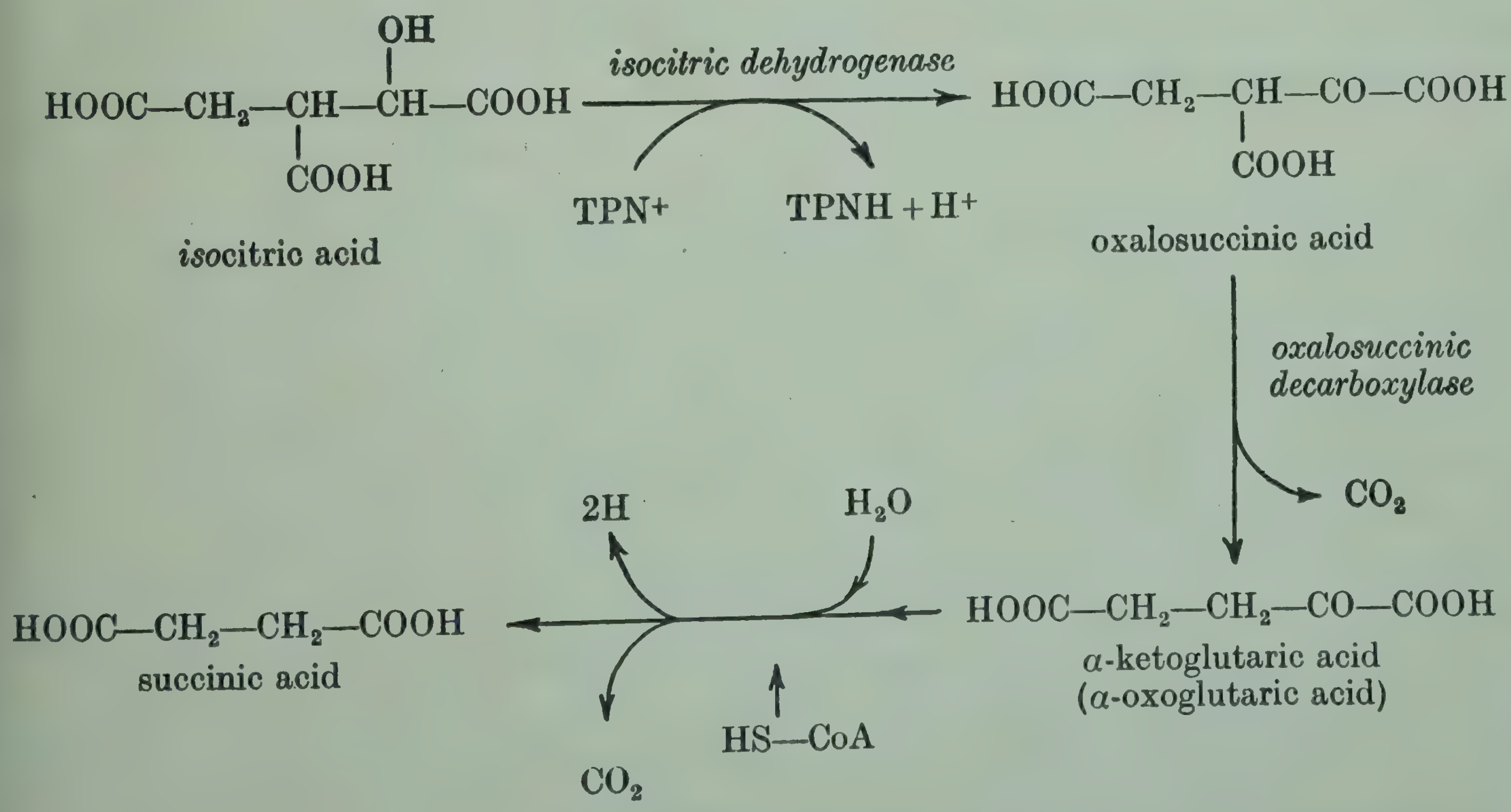
$$HOOC\text{—}CH_2\text{—}\overset{\overset{OH}{|}}{\underset{\underset{COOH}{|}}{C}}\text{—}CH_2\text{—}COOH \xrightarrow[-H_2O]{\textit{aconitase}} HOOC\text{—}CH_2\text{—}\underset{\underset{COOH}{|}}{C}{=}CH\text{—}COOH$$

citric acid *cis*-aconitic acid

$$\xrightarrow[+H_2O]{\textit{aconitase}} HOOC\text{—}CH_2\text{—}\overset{\overset{H}{|}}{\underset{\underset{COOH}{|}}{C}}\text{—}\overset{\overset{OH}{|}}{C}H\text{—}COOH$$

*iso*citric acid

*iso*Citric acid is then oxidized by the removal of two atoms of hydrogen under the influence of *isocitric dehydrogenase* which operates in conjunction with coenzyme II (TPN) (p. 86).



$$HOOC\text{—}CH_2\text{—}\underset{\underset{COOH}{|}}{C}H\text{—}\overset{\overset{OH}{|}}{C}H\text{—}COOH \xrightarrow[TPN^+ \;\to\; TPNH + H^+]{\textit{isocitric dehydrogenase}} HOOC\text{—}CH_2\text{—}\underset{\underset{COOH}{|}}{C}H\text{—}CO\text{—}COOH$$

*iso*citric acid oxalosuccinic acid

$$\xrightarrow[-CO_2]{\textit{oxalosuccinic decarboxylase}} HOOC\text{—}CH_2\text{—}CH_2\text{—}CO\text{—}COOH$$

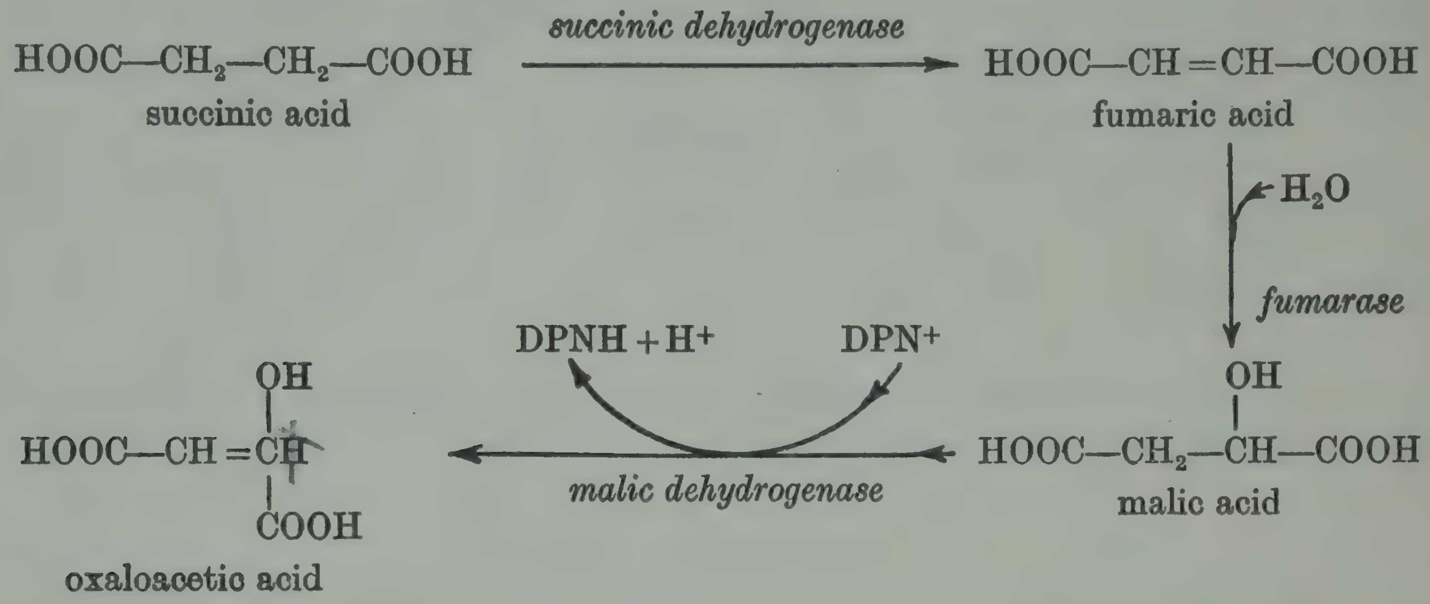
α-ketoglutaric acid (α-oxoglutaric acid)

$$HOOC\text{—}CH_2\text{—}CH_2\text{—}COOH \xleftarrow[CO_2,\; HS\text{—}CoA]{2H,\; H_2O} HOOC\text{—}CH_2\text{—}CH_2\text{—}CO\text{—}COOH$$

succinic acid

The product of this reaction, oxalosuccinic acid, is decarboxylated by loss of carbon dioxide to form α-ketoglutaric acid (α-oxoglutarate) which takes up water, loses carbon dioxide and is oxidized by removal of 2H to form succinic acid. In this complex step coenzyme A plays a part but the details need not be given here.

The succinic acid is then oxidized by removal of 2H, to fumaric acid which takes up water, to form malic acid which in turn is oxidized by *malic dehydrogenase*, acting in conjunction with DPN, to form oxaloacetic acid.



The oxaloacetic acid so formed is now available for reaction with another molecule of acetylcoenzyme A to begin the cycle again.

The whole process is summarized briefly in Fig. 20, 18, from which it will be clear that for each turn of the cycle three molecules of carbon dioxide are liberated (corresponding to the three carbon atoms of pyruvic acid), one in the formation of acetyl coenzyme A and the other two in the citric acid cycle proper. At the same time five pairs of hydrogen atoms are removed from the participating compounds, one in the formation of acetyl coenzyme A and the other four in the cycle proper. These five pairs of hydrogen atoms are passed on (through the appropriate pyridine nucleotide coenzymes as carriers) to the cytochrome system where they are converted to 5 molecules of water, 5 atoms of oxygen being used up in the process. Three of the five water molecules are used up in the cycle itself, and the remaining two are set free so that the oxidation of pyruvic acid by the citric acid cycle fulfils the equation,

$$C_3H_4O_3 + \tfrac{5}{2}O_2 \longrightarrow 3CO_2 + 2H_2O$$

The most important feature of the tricarboxylic acid cycle, however, is that, with the stepwise removal and subsequent oxidation of each pair of hydrogen atoms, the energy so obtained by the breakdown of carbohydrate appears not as heat but in the form of the energy-rich phosphate bonds of ATP. In other words, the oxidation of carbohydrate in its final stages is accompanied by the simultaneous phosphorylation of ADP to ATP, and the combined process is known as *oxidative phosphorylation.* In the oxidation of one molecule of pyruvic acid *at least* 16 energy-rich phosphate bonds are created, that is at least 16 molecules of ATP are produced from ADP. This means that for each 6-carbon unit broken down at least 32 molecules of ATP are produced. Since in glycolysis only 3 molecules of ATP are generated for each 6-carbon unit of glycogen broken down to pyruvic acid, it is clear that the oxidative breakdown of carbohydrate (or pyruvate) in the citric acid cycle constitutes a much more effective mechanism for making use of the energy available in the carbohydrate molecule.

Certain substances such as 2 : 4-dinitrophenol can inhibit phosphorylation without decreasing oxygen utilization. Such substances which disconnect phosphorylation from oxidation are known as *uncoupling agents*, since they uncouple the energy-yielding mechanism (oxidation) from the energy-harnessing system (phosphorylation).

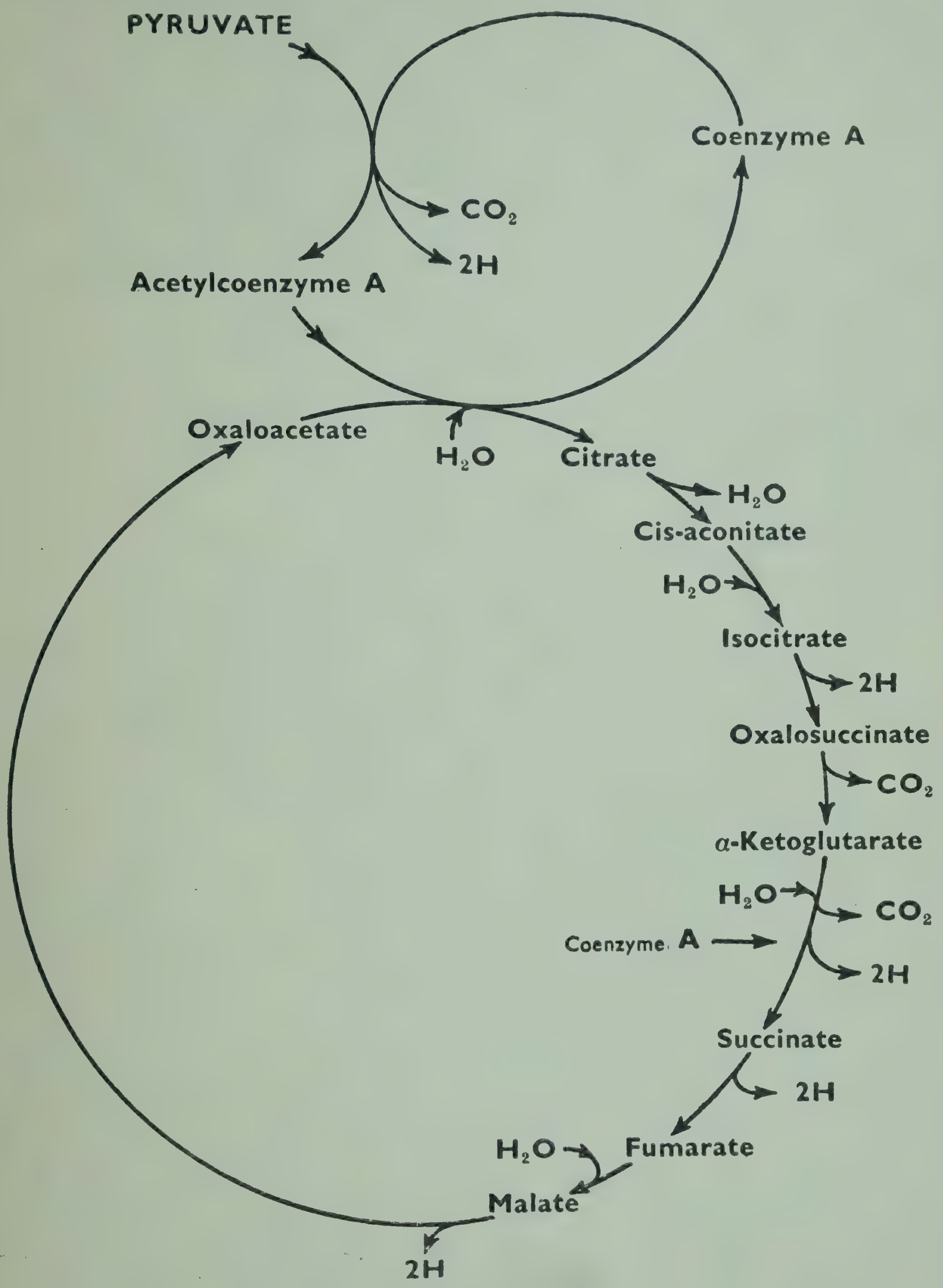
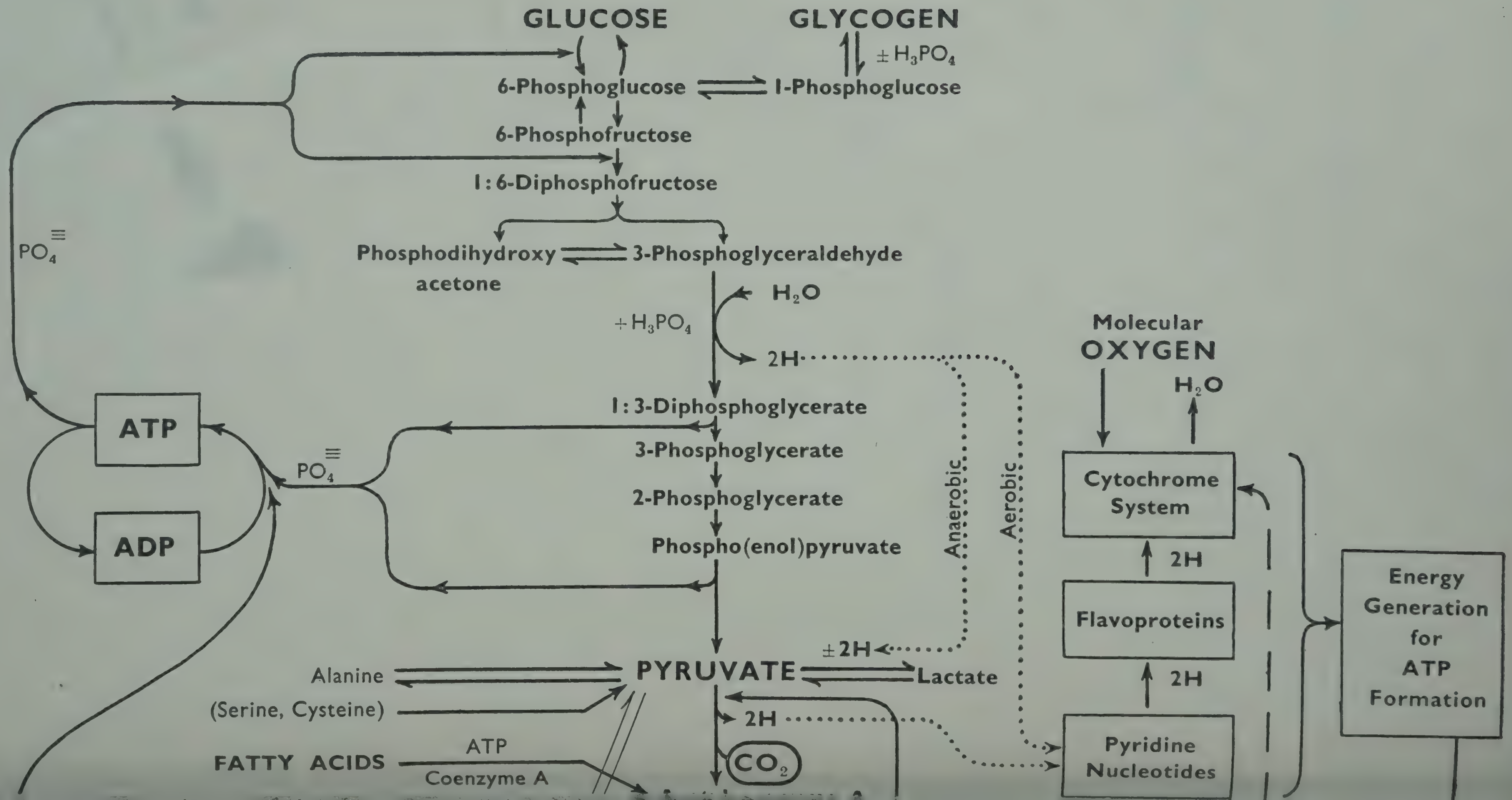
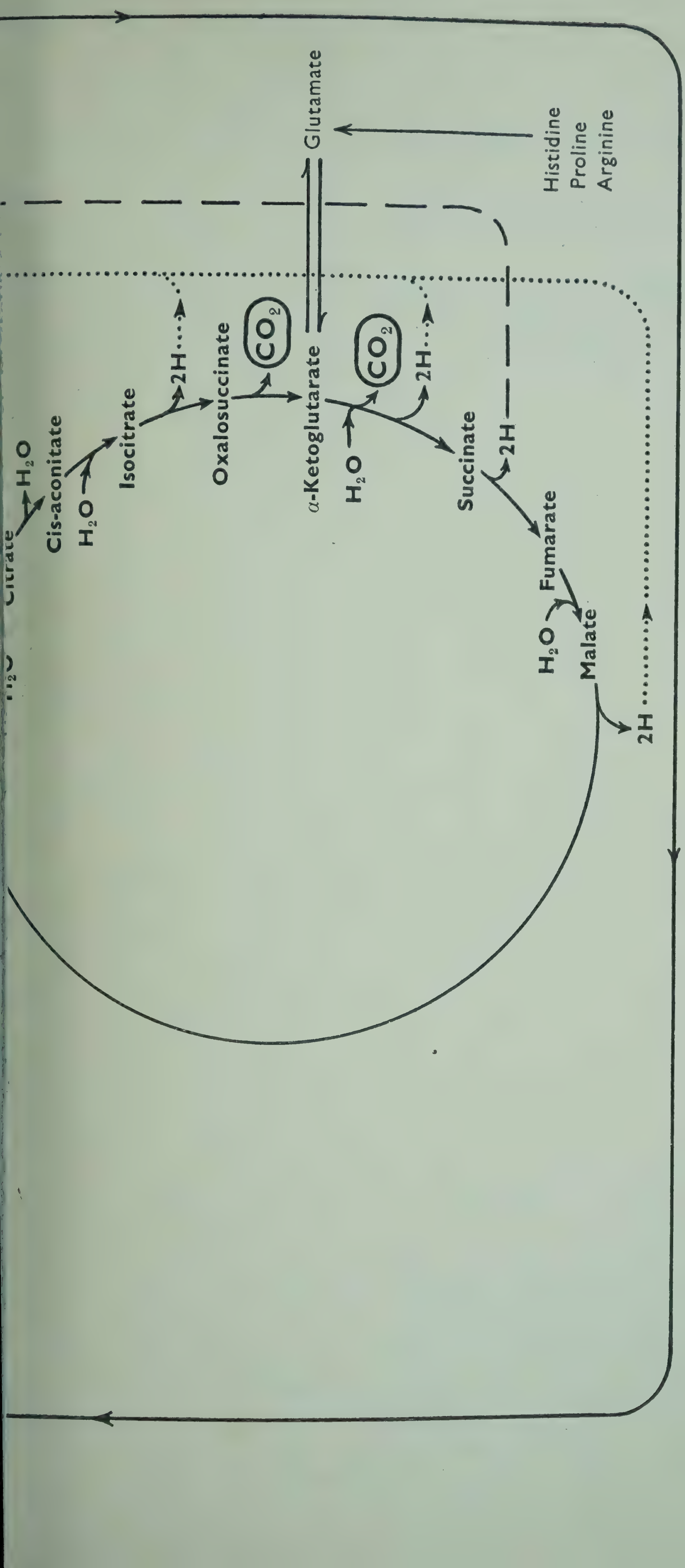

PYRUVATE
Coenzyme A
CO_2
2H
Acetylcoenzyme A
Oxaloacetate
H_2O
Citrate
H_2O
Cis-aconitate
H_2O
Isocitrate
2H
Oxalosuccinate
CO_2
α-Ketoglutarate
H_2O
CO_2
Coenzyme A
2H
Succinate
2H
H_2O
Fumarate
Malate
2H

FIG. 20, 18
The citric acid cycle.

CONCLUSION

A summary of the processes of carbohydrate breakdown is shown in Fig. 20, 19. As will be seen later the terminal stages in the oxidation of fatty acids also involve the citric acid cycle. Moreover, several amino-acids can by oxidative deamination or by transamination be converted into keto acids which are already on the pathway of carbohydrate breakdown. There is therefore a point common to the metabolism of carbohydrates, fats and proteins.


GLUCOSE
GLYCOGEN
± H_3PO_4
6-Phosphoglucose
1-Phosphoglucose
6-Phosphofructose
1:6-Diphosphofructose
$PO_4^{\equiv}$
Phosphodihydroxy acetone
3-Phosphoglyceraldehyde
H_2O
+ H_3PO_4
2H
1:3-Diphosphoglycerate
3-Phosphoglycerate
2-Phosphoglycerate
Phospho(enol)pyruvate
ATP
ADP
$PO_4^{\equiv}$
Anaerobic
Aerobic
Molecular OXYGEN
H_2O
Cytochrome System
2H
Flavoproteins
2H
Pyridine Nucleotides
Energy Generation for ATP Formation
± 2H
PYRUVATE
Lactate
Alanine
(Serine, Cysteine)
FATTY ACIDS
ATP
Coenzyme A
2H
CO_2


Citrate
H_2O
Cis-aconitate
H_2O
H_2O
Isocitrate
2H
Oxalosuccinate
CO_2
α-Ketoglutarate
Glutamate
H_2O
CO_2
2H
Histidine
Proline
Arginine
Succinate
2H
Fumarate
H_2O
Malate
2H

REFERENCES

BALDWIN, E. (1952). *Dynamic Aspects of Biochemistry*, 2nd ed. Cambridge: University Press.

BELL, D. J. (1952). *Introduction to Carbohydrate Biochemistry*, 3rd ed. London: University Tutorial Press.

CORI, G. T. (1952–53). *Harvey Lect.* **48,** 145–166.

SOSKIN, S. & LEVINE, R. (1952). *Carbohydrate Metabolism*, 2nd ed. University of Chicago Press.

YOUNG, F. G. (1951). Experimental approach to diabetes mellitus. *Brit. med. J.* ii, 1167–1170.

YOUNG, F. G. (1953). The growth hormone and diabetes. *Recent Progress in Hormone Research* (Editor G. Pincus), **8,** 471–486.

CHAPTER 21

FAT METABOLISM

THE digestion and absorption of fats have already been discussed. As we have seen, some of the free fatty acids may pass to the liver by way of the portal vein, whereas neutral fat reaches the systemic blood stream in the chyle by way of the thoracic duct. Just as with carbohydrate the ingestion of fat is followed by an increase in the fat content of the blood which may last up to five hours (Fig. 21, 1). Unlike the glucose tolerance test, however, the fat tolerance test has proved of little value in the diagnosis of disease. In health the total plasma lipid concentration (concentration of total ether-soluble matter) varies between the rather wide limits of 400 to 1240 mg. per 100 ml. Physiological increases are found, for example, after meals containing a high percentage of fat, after exercise, in starvation, pregnancy and lactation. It is important to understand that the plasma lipids are very complex and that physiological

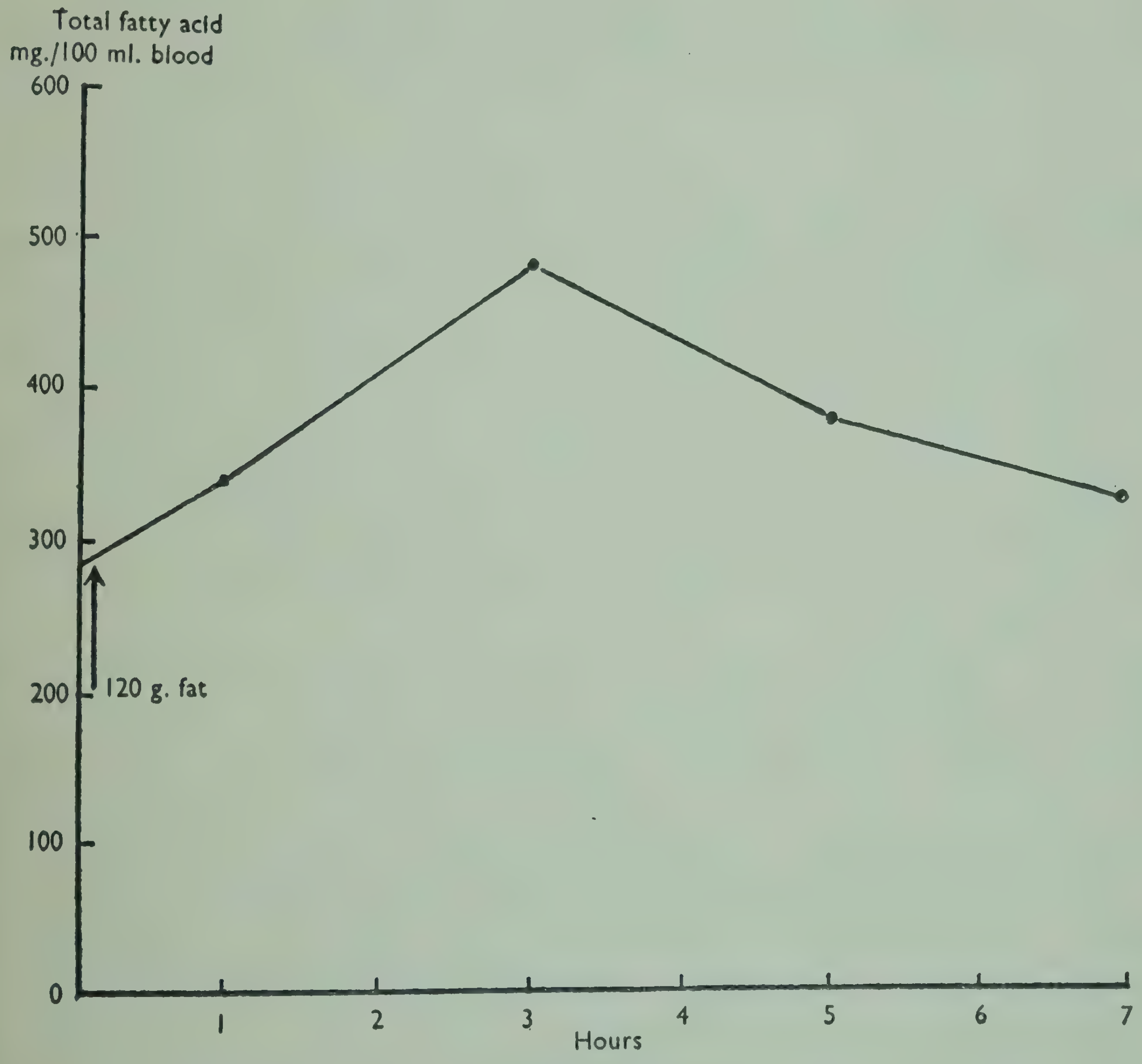


FIG. 21, 1

Variations in blood fatty acid following ingestion of 120 g. of fat by a man aged 41.

variations among the different constituents have not been fully elucidated. Average values for the various forms in which lipid is known to be present in plasma are given in Table 21, 2.

TABLE 21, 2

	mg./100 ml. plasma
Total lipid	589
Neutral fat	154
Total fatty acids	353
Phospholipid fatty acids	130
Cholesterol ester fatty acids	77
Neutral fat fatty acids	146
Total cholesterol	162
Phospholipid	196

Mean values from eight healthy young women under standard conditions. From E. M. Boyd (1933), *J. biol. Chem.* **101**, 323.

Although the lipids are in general insoluble in water, plasma and serum are limpid solutions. The reason for this is that lipids occur in blood plasma as lipid-protein complexes (lipo-proteins) in which the lipid is bound to the protein fraction, chiefly the α- and β-globulins, in an unknown manner. Such lipids cannot be extracted from plasma with ether at ordinary temperatures but, when plasma is frozen with ether to a temperature below −25° C, a large proportion of the lipids is transferred to the ether phase and can be removed. Lipids may also be removed from plasma by treatment with alcohol-ether mixtures at low temperatures (−10° C) or by extraction with fat solvents after denaturation of the proteins with alcohol.

In certain pathological states, for example nephrosis, the lipo-proteins of the blood are often much increased and qualitatively altered.

After absorption fat is either oxidized or stored in the body until required. The liver plays an important role in the metabolism of fat (Fig. 21, 3), but the extent to which fat is either stored or oxidized is in some measure under the control of the endocrine and nervous systems. Some people store but little fat on overfeeding with fat or carbohydrate, while others again lay down fat readily when their dietary intake exceeds their calorie requirements only slightly (see p. 215). The mechanisms underlying these differences are not understood, but genetic factors seem to play an important part in many people.

The chief fat stores of the body are the subcutaneous and retroperitoneal depots. As a rule, fat is stored in a form which is characteristic of the species concerned and ingested fat is converted into this characteristic form during the process of absorption. Thus mutton fat is different from beef fat and both differ from human fat in melting-point, iodine number, and so on, depending on the nature of the fatty acids which are present in the triglycerides. In certain cases, however, an animal may deposit in its depots the fat of another species, as in the classical example of the starved dog which was fed such large amounts of mutton fat that it deposited in its subcutaneous tissues fat characteristic of the sheep.

The fats of the body are not in an inert state even in the fat stores.

If fatty acids labelled with deuterium are fed to animals, nearly half of the depot fat is found to contain deuterium within a week. When palmitic acid labelled with deuterium is fed to rats together with an adequate fat diet, the deuterium can subsequently be found in the depot fats not only in the palmitic acid but also in the stearic, myristic and lauric acids, showing that the carbon chain can be lengthened or shortened. Deuterium can be demonstrated also in the unsaturated acids, oleic acid and palmitoleic acid, showing that desaturation can take place. No deuterium finds its way into the linoleic acid which is one of the essential fatty acids (p. 194). Schoenheimer maintained that lipids are constantly being transported, as fats or phosphatides, from the depots to the organs, such as the liver, where the ester linkages are temporarily broken and fatty acids are subjected to processes of interconversion, synthesis and degradation, in the course of which they become indistinguishably mingled with the fatty acids of dietary origin before either re-entering the ester linkages and returning to the depots or being degraded.

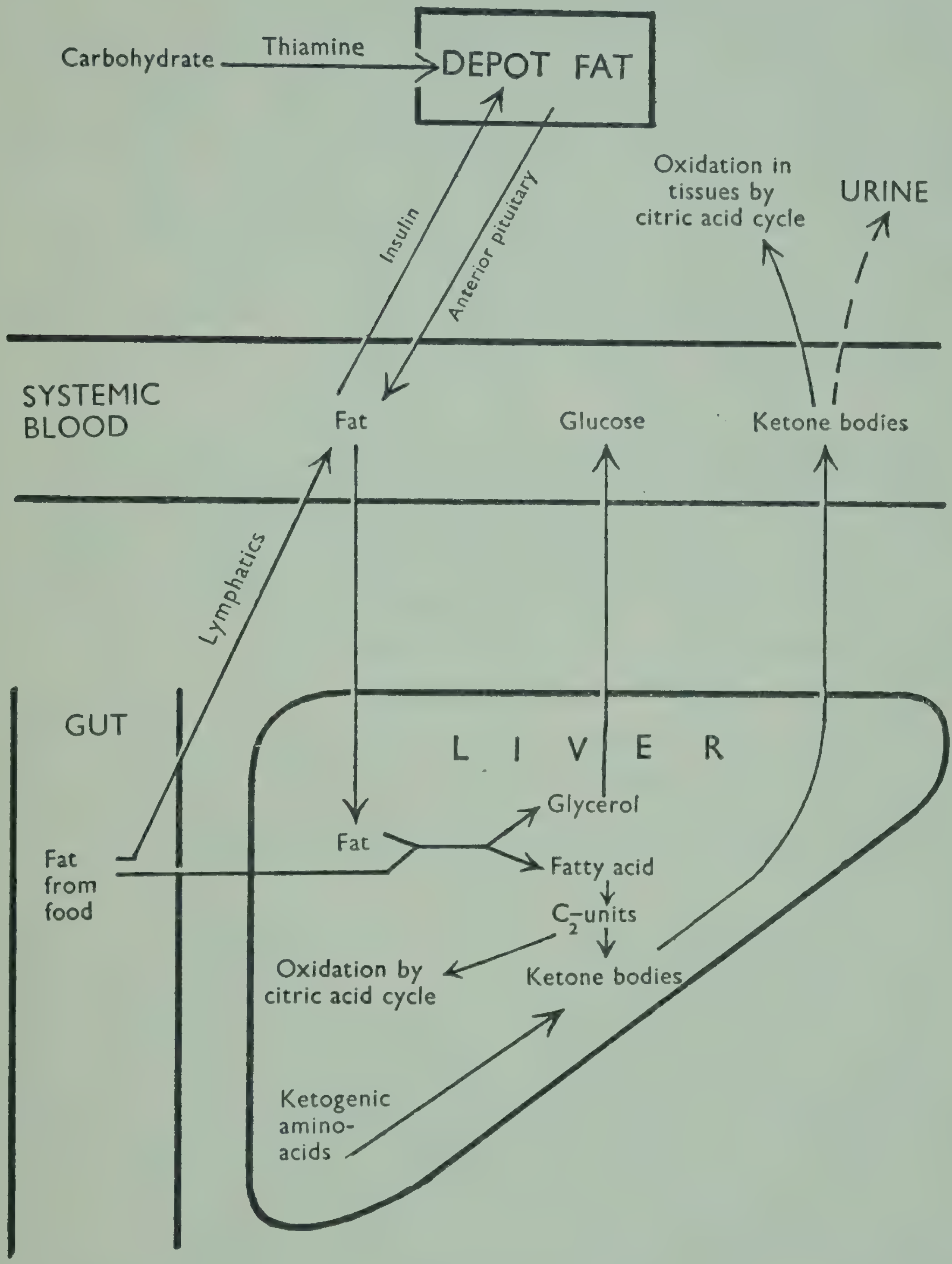

Carbohydrate
Thiamine
DEPOT FAT
Insulin
Anterior pituitary
Oxidation in tissues by citric acid cycle
URINE
SYSTEMIC BLOOD
Fat
Glucose
Ketone bodies
Lymphatics
GUT
L I V E R
Glycerol
Fat
Fatty acid
Fat from food
C₂-units
Oxidation by citric acid cycle
Ketone bodies
Ketogenic amino-acids

FIG. 21, 3

Diagrammatic representation of fat metabolism.

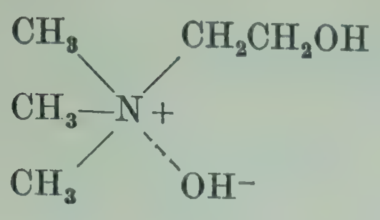
Fat is present in most tissues as well as in the depots. While depot fat is mainly neutral fat (triglycerides), the lipids of the tissues consist both of neutral fat and of phospholipids. The former is mainly stored fat; the latter are present mostly as essential components of the structure of the cells. After prolonged fasting the neutral fat content of tissues becomes greatly depleted, whereas the phospholipid content is much less affected. The phospholipid content of the brain, for example, remains high during starvation, since the cerebral phospholipids do not represent stored material but are essential for the proper functioning of the tissue. It is desirable, however, that the distinction should not be too sharply made between the fats, mainly phospholipids, which are essential components of the cell structure, and which therefore do not disappear on fasting (the 'constant element'), and the storage fat, mainly neutral fat, which can be utilized by the fasting animal (the 'variable element'). Recent work indicates that, especially in the liver, phospholipid may be stored in the form of phospholipid - nucleoprotein complexes, which play a definite part in cellular metabolism but which nevertheless may be used up during fasting.

Some of the lipid material in normal tissues is present in a form which cannot be stained histologically by the usual fat stains, probably because it is present in the form of phospholipid-protein complexes. In the process of *fatty degeneration* which is found histologically for example with anæmia and infection, such material becomes 'unmasked' and can readily be stained although the total fat content of the tissue need not be increased. On the other hand, poisoning with arsenic, phosphorus, chloroform, carbon tetrachloride and some other drugs, and many infective processes are associated with a greatly increased fat content of the liver; the mechanism is not clearly understood but, at any rate in animals, the degree of increase of fat content is dependent on the nutritional state.

In the cytoplasm of the cell much of the lipid is present in the form of lipo-protein complexes and cannot be demonstrated microscopically unless the tissue has previously been treated with proteolytic enzymes, or with reagents such as alcohol which destroy the lipid-protein links. In the mitochondria, for example, phospholipid is present in the form of phospholipid-ribonucleoprotein complexes, and similar complexes are present in the submicroscopic particles of the liver cell.

As we have already mentioned, the liver plays a special role in the metabolism of fat. The normal liver contains about 4 per cent. of lipid material of which only one-quarter is neutral fat while three-quarters are phospholipid. The amount of neutral fat rises markedly during starvation as fat passes from the depots to the liver to be oxidized. As the stores become exhausted the fat content of the liver falls.

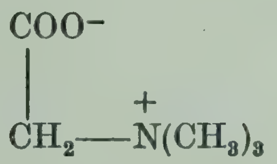
The lipid content of the liver may, however, be increased in other circumstances. When large amounts of neutral fat (or of cholesterol) are fed to rats, the fat content of the liver is greatly increased. It was shown by Best and his colleagues in Toronto that such abnormal fat deposition could be abolished by the inclusion of lecithin in the diet, and later the active constituent of the lecithin was found to be choline. The addition of as little as 1 mg. of choline per rat per day to the food is sufficient to prevent the excessive deposition of fat in the liver.



choline

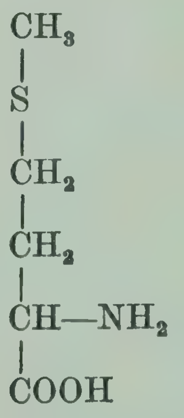
When choline is absent from the diet not only does fat tend to be deposited in the liver cells but a progressive fibrosis of the liver (cirrhosis) ultimately develops. The fatty change can be corrected by supplementing the diet with choline (and in other ways), but the fibrosis is irreversible and may lead to the death of the animal. Although a connexion has clearly been established experimentally between such agents as choline and cirrhosis of the liver, the relationship of these factors to disease in man is not yet clear. In addition to these lesions in the liver, when choline is absent from the diet pathological changes may be found in other organs. Hæmorrhagic degeneration is found in the kidneys of young male rats on a choline-free diet. In young fowls the condition of *perosis*, in which the tendo calcaneus slips and the hock joint is distorted, or dislocated, is caused by the absence of choline and manganese from the diet.

Choline is not the only substance which can prevent the deposition of abnormally large amounts of fat in the liver. Betaine exercises a similar effect and so does casein. It has been shown that the action of casein is due to the amino-acid methionine which it contain~~s:~~



$$\begin{array}{l} COO^- \\ | \\ CH_2\text{—}\overset{+}{N}(CH_3)_3 \end{array}$$

betaine



$$\begin{array}{l} CH_3 \\ | \\ S \\ | \\ CH_2 \\ | \\ CH_2 \\ | \\ CH\text{—}NH_2 \\ | \\ COOH \end{array}$$

methionine

Both betaine and methionine possess labile methyl groups which are used in the *in vivo* synthesis of choline by the process of *transmethylation* (p. 347).

Such substances as choline, betaine and methionine are termed *lipotropic* ~~*factors*~~ and are said to exercise a *lipotropic action*. A diet free from choline and other lipotropic factors is said to be *alipotropic*. On the other hand, it has been shown that the amino-acid *cystine*, when added to the diet, causes a large increase in liver fat of rats on a low-casein high-fat diet. Cysteine and homocysteine exercise a similar effect. The term *antilipotropic* has been used to describe this action.

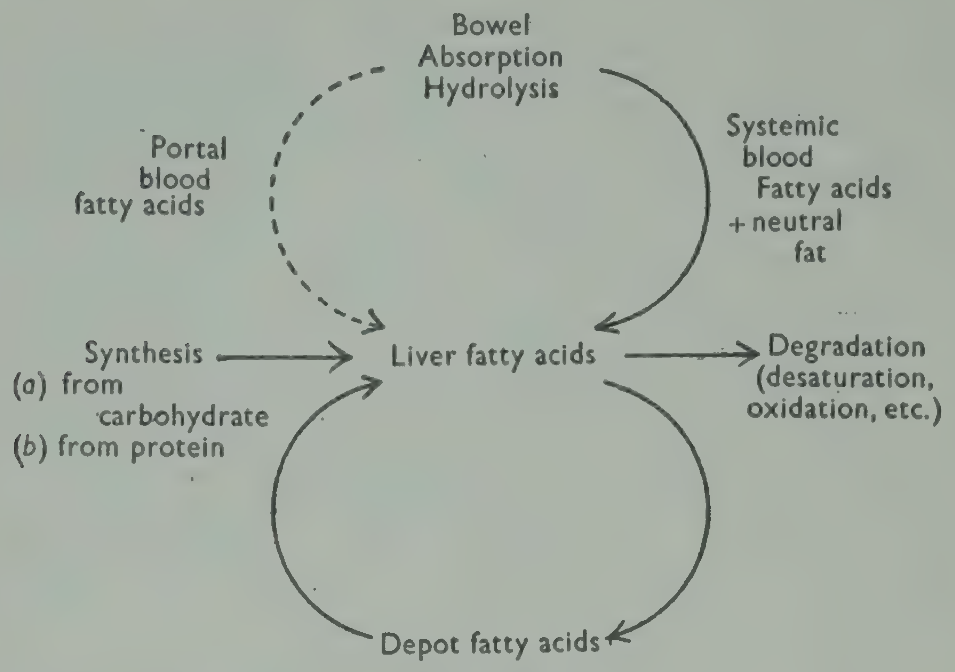
It is probable that choline produces its lipotropic action by stimulating phospholipid interchange between the liver and the other tissues or the fat depots. This view is supported by experiments with isotopes.

Fatty livers may also be produced by a number of other methods. The fatty livers which follow administration of anterior pituitary extract to animals are probably due to transference of fatty acids from the fat depots to the liver; the administration of choline has no effect upon the condition.

SOURCES OF BODY FAT

There are three possible sources of fat in the body (Fig. 21, 4):

(1) The main part of the body fat is derived from the fat taken in the diet.



Bowel
Absorption
Hydrolysis
Portal blood fatty acids
Systemic blood Fatty acids + neutral fat
Synthesis (a) from carbohydrate (b) from protein
Liver fatty acids
Degradation (desaturation, oxidation, etc.)
Depot fatty acids

FIG. 21, 4

(2) Fat may be synthesized from carbohydrate. The body is able to store large amounts of fat but only limited amounts of carbohydrate; thus any excess carbohydrate from the diet, which is not immediately oxidized or stored as glycogen, is converted to fat. The fattening properties of a diet rich in sugar and starch are well known. For the conversion of carbohydrate to fat, thiamine (vitamin B_1) is essential and probably also pantothenic acid as a constituent of coenzyme A. Insulin is also required.

(3) Fat may be produced from protein, although this process probably does not occur to any great extent under ordinary conditions. Longenecker, for example, found that rats on a diet rich in casein and free from fat and carbohydrate showed a large increase in body fat. The mechanism for the synthesis of fatty acid chains is discussed later on p. 330.

In spite of the ability of the tissues to synthesize fat from non-lipid precursors, some of the more unsaturated fatty acids which are essential for the orderly function of the tissues cannot be synthesized and must be obtained from the food (p. 194).

THE OXIDATION OF FATS

The liver normally contains a store of fat and phospholipid, but it also receives from the food or from the fat depots a continuous supply of fat with which it deals by the process of oxidation. It is possible that the fats are first converted to phospholipids before being transported in the blood stream from the depots to the liver, and before being oxidized, but this is not yet certain. In the liver, however, whether triglycerides or phospholipids are to be metabolized, the essential part of the process consists in the oxidation of the long-chain fatty acids. The glycerol portion of the fat molecule is probably converted to glycogen and is utilized after the manner of carbohydrates, although some may be metabolized by way of phosphodihydroxy acetone (p. 312).

Several theories have been proposed to explain the mechanism of the oxidation of the fatty acid chains. The classical theory of *β-oxidation* was the outcome of the work of Knoop. According to this mechanism, fatty acid chains are oxidized by the removal of two carbon atoms at a time. The carbon atom in the β-position to the carboxyl group is presumed to be attacked with the formation of the corresponding β-keto-acid, following which the two terminal carbon atoms are split off as acetic acid. A new carboxyl group (—COOH) is formed at the site of the keto (=CO) grouping so that a fatty acid remains with two carbon atoms fewer than the original. Again the new β-carbon atom is attacked and two more carbon atoms are split off. In this way the fatty acid is broken down by the removal of two carbon atoms at a time, until finally the stage of aceto-acetic acid (β-keto-butyric acid) is reached.

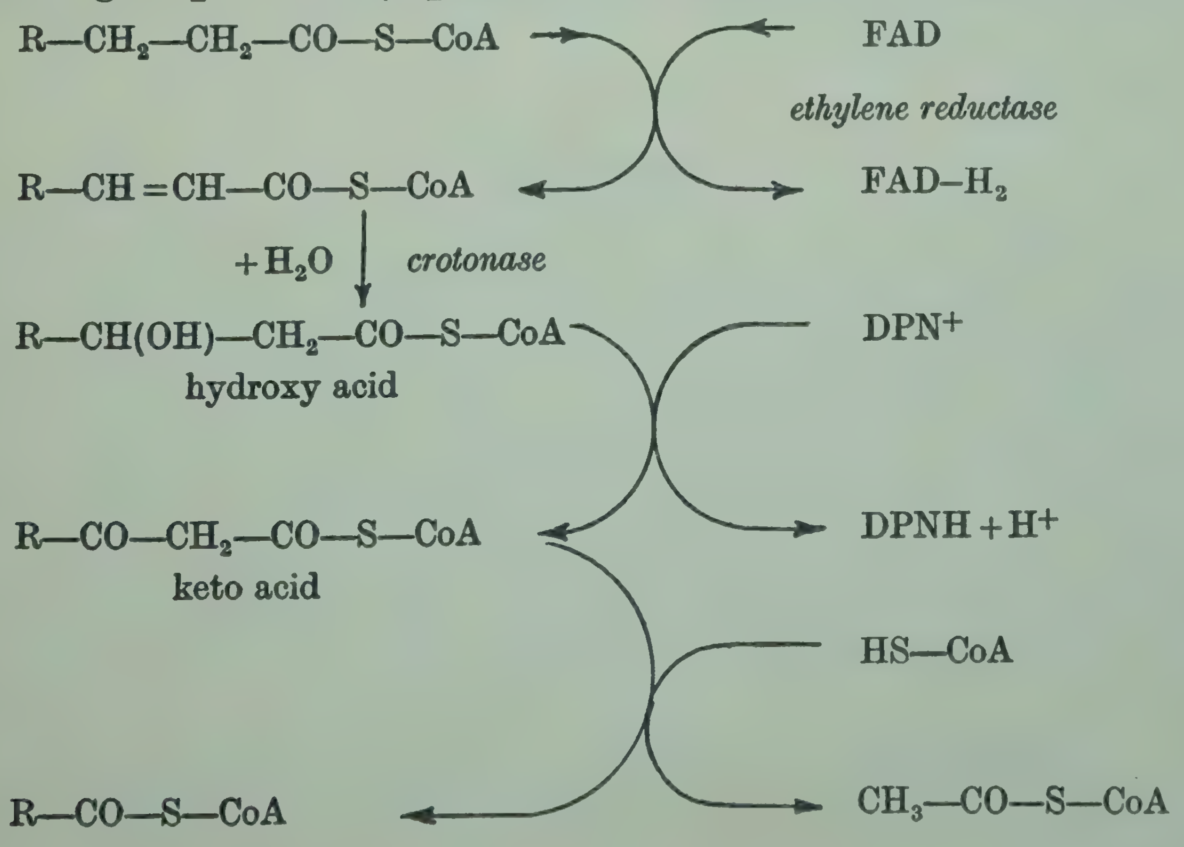
Although this theory has had to be greatly modified in the light of recent discoveries, the essential concept of the stepwise removal of 2-carbon units remains.

According to the modern view of Lynen and others the oxidation of fatty acids involves the essential participation of coenzyme A (HS—CoA) which has already been mentioned on p. 62 and p. 316.

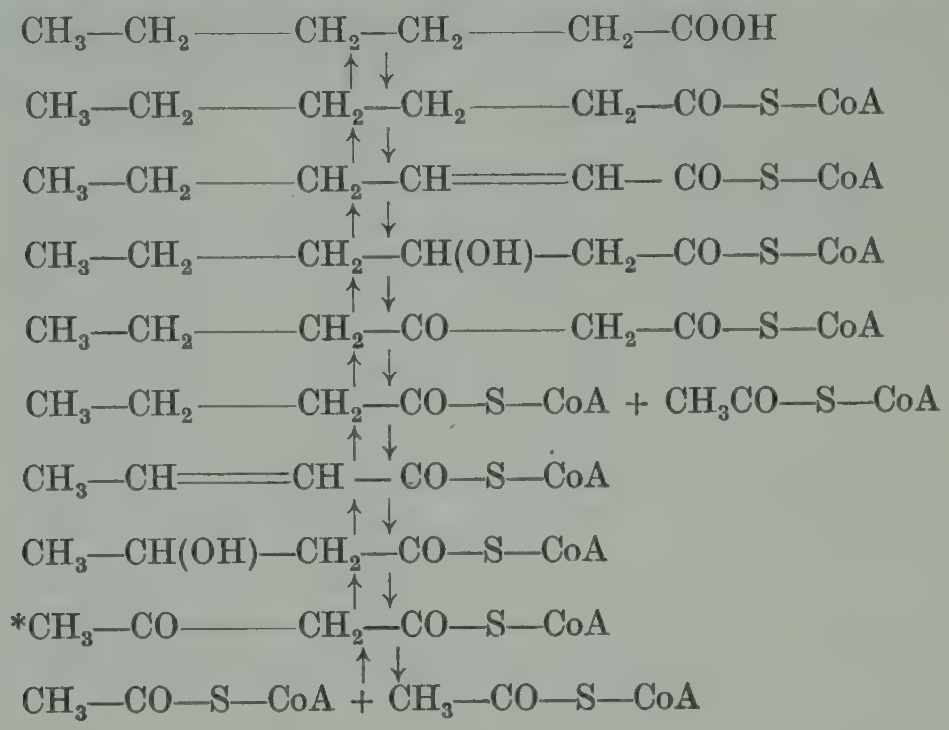
In the initial stages of the breakdown of a fatty acid chain, the fatty acid combines with coenzyme A in a rather complex reaction in which ATP plays an essential part. If we represent the fatty acid as RCH_2CH_2COOH where R may contain any odd number of carbon atoms, the reaction may be shown thus:—

$$RCH_2CH_2COOH + HS—CoA \xrightarrow{+ATP} RCH_2CH_2CO—S—CoA + H_2O$$

The compound so formed is then oxidized by removal of two atoms of hydrogen under the influence of the copper-containing enzyme *ethylene reductase* (*acyl-CoA-dehydrogenase*) with its coenzyme flavin-adenine dinucleotide (FAD). The product takes up water under the influence of the enzyme *crotonase* to form the hydroxy acid in which the hydroxyl group is attached to the carbon atom in the β-position to the original carboxyl carbon. This hydroxy acid is then oxidized by *β-ketoreductase* (*β-hydroxy-acyl-CoA-dehydrogenase*) acting along with DPN to yield the keto-acid with the keto group in the β-position to the original carboxyl carbon.



The keto acid finally reacts with another molecule of coenzyme A forming acetyl coenzyme A and the coenzyme A derivative of the fatty acid containing two carbon atoms fewer than the original. This derivative then repeats the process. The end result of this process is that the fatty acid chains (which all contain even numbers of carbon atoms) are degraded two carbon atoms at a time, the carbon atom being split off as acetyl coenzyme A. The breakdown of caproic acid (six carbon atoms) could be represented as follows, but the process is similar for all long chain fatty acids.



$$\begin{array}{l}
\mathrm{CH_3{-}CH_2{-}CH_2{-}CH_2{-}CH_2{-}COOH} \\
\qquad\quad \uparrow\downarrow \\
\mathrm{CH_3{-}CH_2{-}CH_2{-}CH_2{-}CH_2{-}CO{-}S{-}CoA} \\
\qquad\quad \uparrow\downarrow \\
\mathrm{CH_3{-}CH_2{-}CH_2{-}CH{=}CH{-}CO{-}S{-}CoA} \\
\qquad\quad \uparrow\downarrow \\
\mathrm{CH_3{-}CH_2{-}CH_2{-}CH(OH){-}CH_2{-}CO{-}S{-}CoA} \\
\qquad\quad \uparrow\downarrow \\
\mathrm{CH_3{-}CH_2{-}CH_2{-}CO{-}CH_2{-}CO{-}S{-}CoA} \\
\qquad\quad \uparrow\downarrow \\
\mathrm{CH_3{-}CH_2{-}CH_2{-}CO{-}S{-}CoA + CH_3CO{-}S{-}CoA} \\
\qquad\quad \uparrow\downarrow \\
\mathrm{CH_3{-}CH{=}CH{-}CO{-}S{-}CoA} \\
\qquad\quad \uparrow\downarrow \\
\mathrm{CH_3{-}CH(OH){-}CH_2{-}CO{-}S{-}CoA} \\
\qquad\quad \uparrow\downarrow \\
{}^{*}\mathrm{CH_3{-}CO{-}CH_2{-}CO{-}S{-}CoA} \\
\qquad\quad \uparrow\downarrow \\
\mathrm{CH_3{-}CO{-}S{-}CoA + CH_3{-}CO{-}S{-}CoA}
\end{array}$$

These reactions, at least as far as the stage of acetoacetyl coenzyme A (marked with an asterisk *), take place in the liver.

It should be noted that all these steps are reversible and are in fact reversed when fatty acids are being synthesized in the body from 2-carbon units.

From what has been said above it is clear that fatty acid breakdown results in the formation of one molecule of acetyl coenzyme A for every two carbon atoms in the fatty acid chain. This acetyl coenzyme A now enters the citric acid cycle (Fig. 21, 5), reacting with oxaloacetic acid to form citric acid in the same manner as was described in the preceding chapter. The final pathway for the oxidation of fatty acids to carbon dioxide and water is therefore common to both fat and carbohydrate metabolism (Fig. 20, 19).

KETOSIS

Since acetyl coenzyme A is normally consumed by entering the citric acid cycle in combination with oxaloacetate, the level of fatty acid oxidation is therefore determined to a large extent by the rate at which oxaloacetate is made available. If this rate falls off through the depression of carbohydrate metabolism, in diabetes mellitus or in starvation, then acetyl coenzyme A cannot be metabolized completely. Its immediate precursor in the terminal oxidation stage, acetoacetyl coenzyme A (marked with an asterisk * in the scheme shown above), therefore accumulates and is hydrolysed by a *deacylase* found only in liver to yield acetoacetic acid.

$$\mathrm{CH_3COCH_2CO{-}S{-}CoA} \xrightarrow[\textit{deacylase}]{+\mathrm{H_2O}} \mathrm{CH_3COCH_2COOH + HS{-}CoA}$$

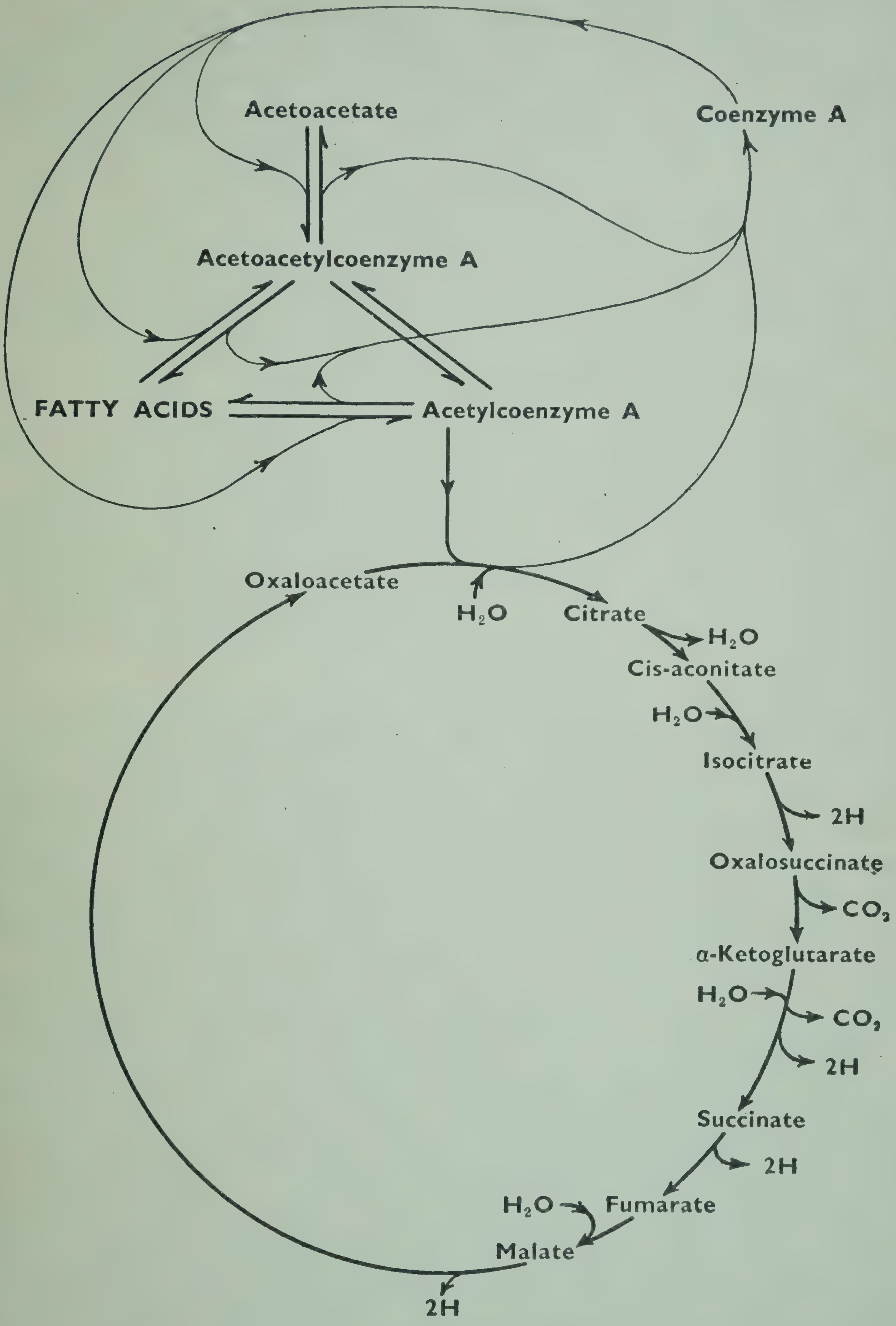
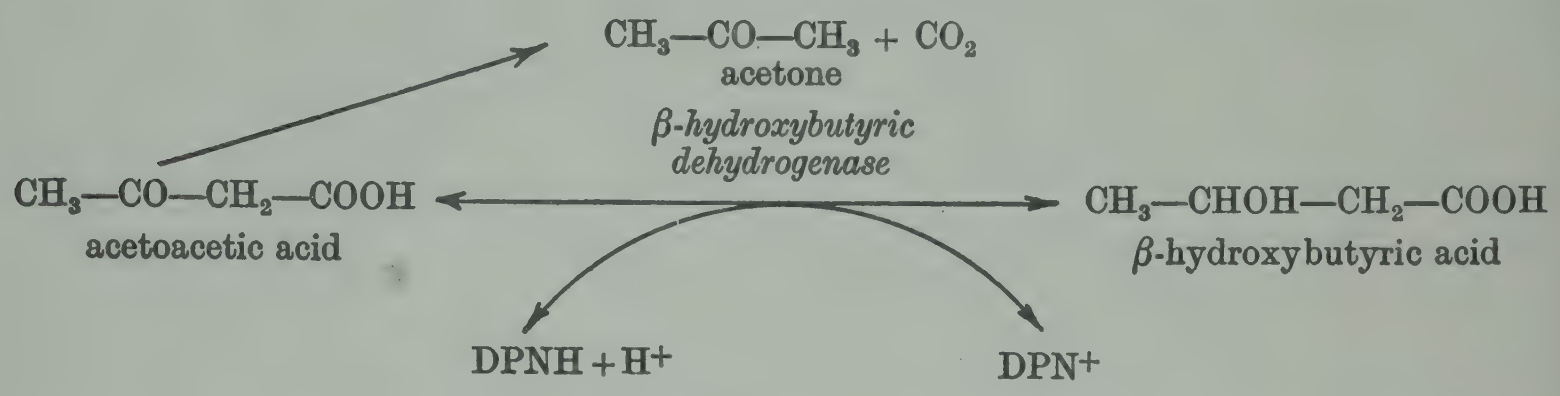

Acetoacetate
Coenzyme A
Acetoacetylcoenzyme A
FATTY ACIDS
Acetylcoenzyme A
Oxaloacetate
H_2O
Citrate
H_2O
Cis-aconitate
H_2O
Isocitrate
2H
Oxalosuccinate
CO_2
α-Ketoglutarate
H_2O
CO_2
2H
Succinate
2H
H_2O
Fumarate
Malate
2H

Fig. 21, 5
The entry of fatty acids into the tricarboxylic acid cycle.

Appreciable amounts of acetoacetic acid are also formed by the condensation of 2-carbon fragments, that is by the condensation of two molecules of acetyl coenzyme A, (Fig. 21, 5).

The acetoacetic acid, whatever be the route by which it is formed, is readily decarboxylated under physiological conditions to *acetone*, and it is also readily reduced to β-hydroxybutyric acid under the influence of the enzyme *β-hydroxybutyric dehydrogenase* acting in conjunction with DPN.



These three compounds are known as the *ketone bodies* or *acetone bodies.* They are produced in the process of *ketogenesis* which occurs mainly in the liver and they are disposed of by oxidation through the citric acid cycle in the process of *ketolysis* which occurs mainly in the *extrahepatic* tissues such as *kidney* and muscle. In this process they are activated by coenzyme A and ATP to form such compounds as acetoacetyl coenzyme A which can act as the immediate precursor of acetyl coenzyme A.

The term *ketosis* is applied to the condition in which ketone bodies accumulate in the blood (*ketonœmia*) and appear in the urine (*ketonuria*). Ketosis occurs most commonly in starvation and clinical or experimental diabetes mellitus, when the stores of liver glycogen are low. The ketosis of diabetes mellitus is accompanied by profound metabolic disturbances which lead to gradually deepening coma and finally to death. The odour of acetone can readily be appreciated in the patient's breath and urine and amounts of β-hydroxybutyric acid exceeding 200 mg. may be excreted in the urine in twenty-four hours instead of the normal 5 to 10 mg.

From what has already been said, it will be clear that an accumulation of ketone bodies may be due either to an abnormally high rate of formation or to an abnormally low rate of disposal. The capacity for disposal is limited but is not normally exceeded so long as fat is not the only substance being oxidized. In the tissues, fatty acids and carbohydrate compete for the available oxygen but in the normal organism carbohydrate is preferentially oxidized. In starvation carbohydrate supplies are low and in diabetes they cannot be properly utilized; the result is that fatty acid oxidation may be increased to such a level that the production of ketone bodies outstrips the limited capacity of the body to dispose of them.

Ketogenesis can be checked by the administration of carbohydrate provided that the carbohydrate can be properly utilized, that is in normal subjects and in diabetics receiving insulin. Carbohydrate is therefore an *antiketogenic* substance and its action is termed *antiketogenesis.* Similarly the glycerol portion of the fat molecule is antiketogenic. The deamination of amino-acids gives rise to keto-acids (p. 345), but not all these are ketogenic. They may be divided (p. 345) into those that are metabolized by way of aceto-acetic acid (ketogenic amino-acids) and those by way of carbohydrate derivatives (antiketogenic amino-acids). It is probable that most, if not all, antiketogenic substances act as fuel which is metabolized in preference to fat. They, therefore, cut down fatty acid oxidation and cause a decrease in aceto-acetic acid production (fat-sparing action).

If the carbohydrate content of the diet is adequate, and if carbohydrate can be utilized, fatty acid oxidation does not proceed at such a rate that ketogenesis exceeds ketolysis. In the past, diets for diabetic patients were constructed with great care so that the ratio of ketogenic to antiketogenic materials did not exceed 1·5 to 1. The use of insulin makes such a procedure only occasionally necessary.

It has been known for a long time that the production of ketone bodies is to some extent under the control of the endocrine system and that extracts of the anterior lobe of the pituitary gland can bring about a rise in the ketone bodies of the blood even in the presence of adequate carbohydrate supplies. This rise can be prevented, however, by the administration of insulin prior to the injection of anterior pituitary extract. Both insulin and anterior pituitary hormone exert their action largely in the liver, since neither is effective in animals from which the liver has been removed. More recently H. B. Stewart and F. G. Young have isolated from muscle tissue a crystalline lipid-like substance which causes the excretion of ketone bodies in the urine when it is administered to rats, normal or alloxan-diabetic, on a high fat diet containing protein in the form of casein.

THE METABOLISM OF CHOLESTEROL

Ingested cholesterol is absorbed along with other lipids. Many other sterols, however, including most plant sterols, are not absorbed from the gut and are not utilized by the body. Cholesterol is normally present in blood to the extent of 100 to 200 mg. per 100 ml., being equally divided between the cells and the plasma. In the cells cholesterol occurs mainly in the free form, whereas in the plasma about 75 per cent. is found in the form of cholesterol esters.

In man, the level of plasma cholesterol normally bears little relation to the amount consumed in the diet, although if a diet containing little fat or cholesterol is consumed for several months there is a fall in concentration of cholesterol in the plasma, especially if this was elevated at the beginning of the period. In the chicken and the rabbit, however, greatly elevated blood cholesterol concentrations have been produced by adding cholesterol to the diet and very high levels are found in certain diseases of man including diabetes mellitus and myxœdema. The experimental animals and patients suffering from these disorders are frequently found to have arterial disease (*atherosclerosis*). This and other evidence suggests a relationship between cholesterol and fat metabolism and arterial disease.

Little is known of the metabolism of cholesterol but it is probable that this substance may be the source from which the body synthesizes other steroids such as sex hormones and bile acids. Cholesterol, hitherto considered to be chemically inert and metabolically inactive, is now known to be an essential constituent of the cell and an active metabolite. Experiments with isotopes have shown that cholesterol can be synthesized in the body from C_2-units and that there is a surprisingly rapid interchange between the cholesterol of the plasma and that of the liver and other tissues. It is apparently synthesized by the tissues from simple carbon compounds and it has been estimated that human beings synthesize as much as 1·5 to 2·0 g. of cholesterol per day, an amount much greater than that normally ingested.

REFERENCES

COOK, R. P. (1942–43). Cholesterol metabolism. *Nutr. Abst. Rev.* **12,** 1–11.

DEUEL, H. J. (1955). *The Lipids,* Vol. 2. London: Interscience.

Lipid metabolism (1952). *Biochem. Soc. Symp.* No. 9.

LIPMANN, F. (1954). Development of the acetylation problem. *Science,* **120,** 855-900.

LYNEN, F. (1952–53). *Harvey Lect.* **47,** 210–224.

CHAPTER 22

PROTEIN METABOLISM

THE ultimate products of protein digestion are the amino-acids. Amino-acids are absorbed from the intestine into the portal blood stream by which they are conveyed to the liver. Here the bulk of them is removed from the blood so that the level of amino-acid nitrogen in the systemic circulation rises but little over the normal value of about 4 mg. per 100 ml., although the blood urea nitrogen is somewhat increased following a protein meal (Fig. 22, 1).

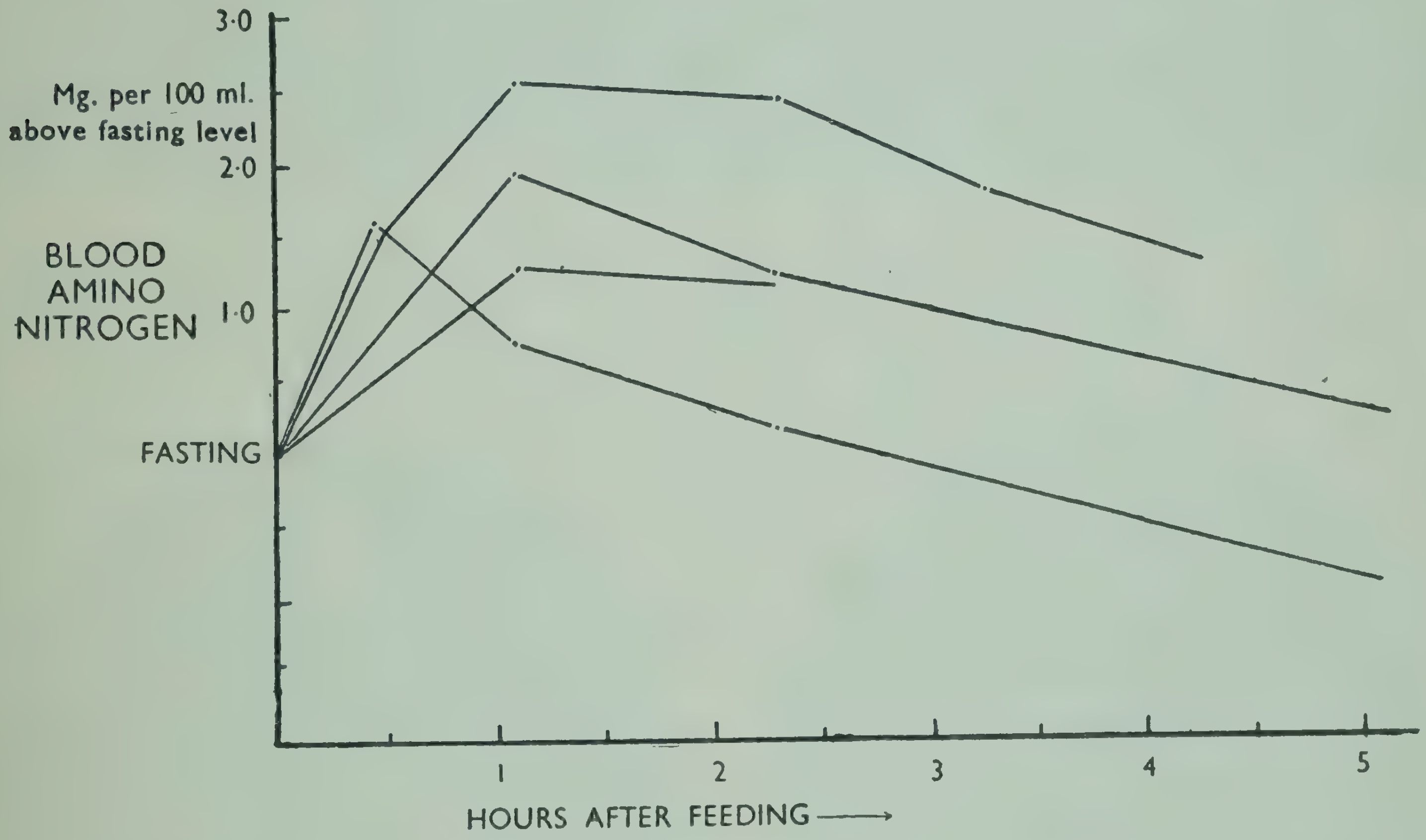


FIG. 22, 1

Amino nitrogen absorption curves following ingestion of casein by four normal infants. The ordinate gives blood amino nitrogen in mg. per 100 ml. above the fasting value (C. D. West, J. L. Wilson, & R. Eyles (1946), *Amer. J. Dis. Child.* **72**, 251).

The injection of a solution containing a mixture of amino-acids directly into the systemic circulation is followed by a rapid disappearance of amino-acids from the blood. They are very quickly taken up by the kidney, by the muscles, and especially by the liver. As the amino-acids are removed from the blood by the liver an amount of urea corresponding to their nitrogen is liberated into the blood by a mechanism which will be considered in detail on p. 343.

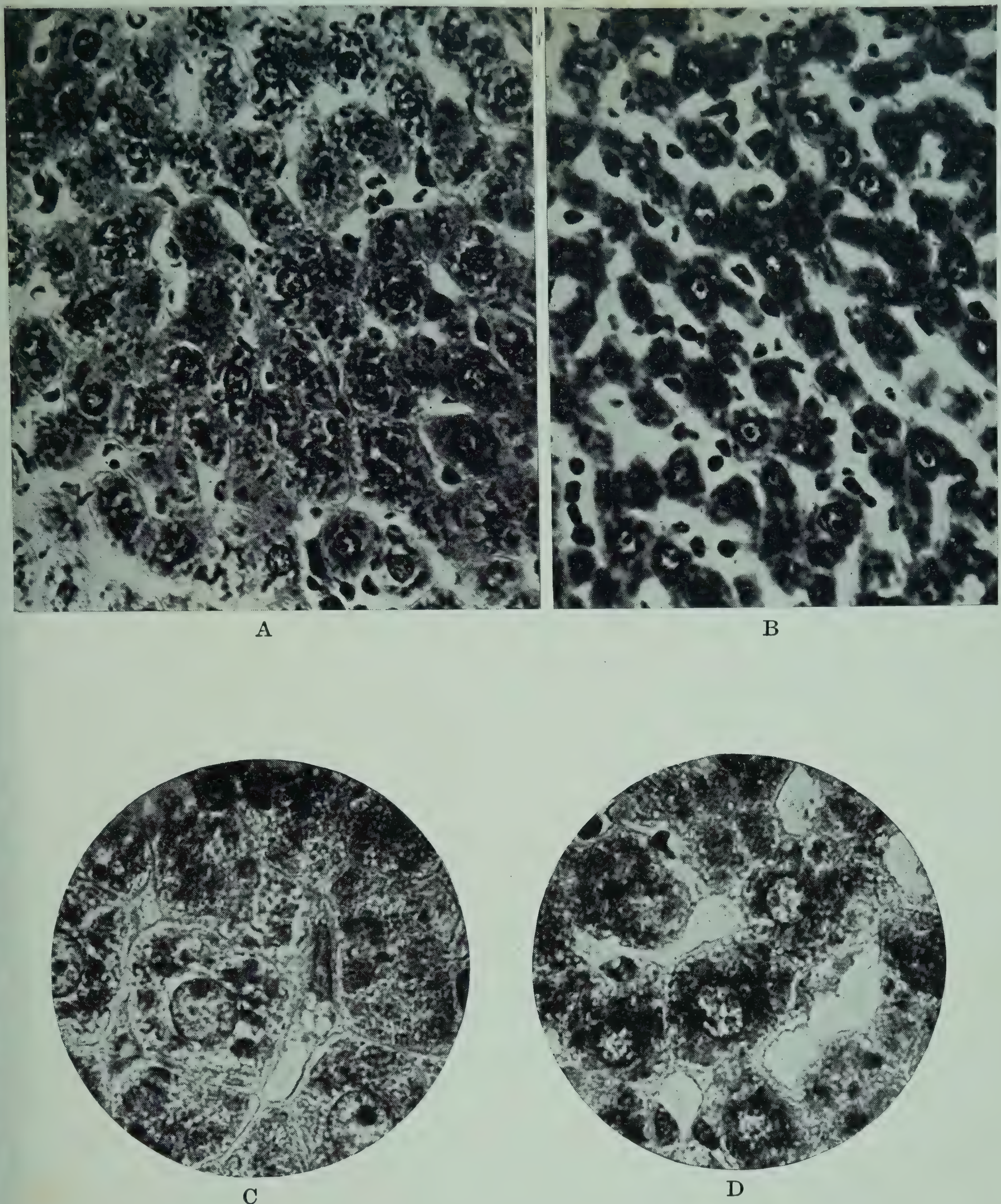
In contrast to the well-known ability of the body to store large amounts of both carbohydrate (p. 297) and fat (p. 324) the tissues have relatively little capacity for storing either amino-acids or protein. However, after a rich protein meal temporary storage of amino-acids takes place both in the liver and in other tissues, and is followed by an increase of the protein in the tissue. The amino-acids needed to form tissue protein are picked out, as it were, from the blood and built up into the proteins characteristic of each tissue. In the adult such tissue formation is required to replace that lost in the process of normal wear and tear but in growing children, and in adults recovering from wasting illnesses, protein synthesis takes place to a very much greater extent since new tissue is being formed. In the liver the protein synthesized from amino-acids (following a protein meal) is stored in that organ partly in the form of granules in the cytoplasm of the cells (Fig. 22, 2). This represents a labile store of protein which is used up rapidly in the early stages of starvation or in other conditions of deficient protein intake, or in helping to make good the loss of plasma protein following, for example, a severe hæmorrhage or loss of blood at surgical operation. In starvation, or when the intake of protein is impaired (for example after certain surgical operations), the liver loses protein to a greater extent than does any other tissue. In recent years evidence has accumulated that the liver, when depleted of protein in these ways, is more susceptible to injury by such agents as phosphorus, arsenic or alcohol which may reach it in the blood; in the modern treatment of liver disease a diet containing a relatively high proportion of protein therefore has an important place.

In certain circumstances, as in some cases of disease or injury to the œsophagus or stomach, where feeding by the mouth is impossible or impracticable, the nitrogen and amino-acid requirements of the body may be met by the administration of amino-acids directly into the blood stream by continuous intravenous infusion. Protein digests or hydrolysates (for example of casein), suitably prepared for intravenous administration, are used for this purpose. Nitrogen equilibrium (p. 188) may thus be maintained for some weeks without protein ingestion, provided that the mixtures employed contain all the essential amino-acids (pp. 185 and 351).

Some of the plasma proteins (p. 369)—albumin, fibrinogen and prothrombin—are synthesized in the liver, whence they are liberated into the blood stream. Injury or disease of the liver may therefore be accompanied by a lowered level of these constituents in the blood (Fig. 22, 3). Experiments with isotopic nitrogen have proved that there is a dynamic equilibrium between the plasma proteins and those of the liver and of other tissues generally. When the level of the plasma proteins is raised or lowered experimentally there is a corresponding rise or fall in the protein content of the liver and of the tissues.

During metabolism of amino-acids the heat produced by the body is increased. This increase of heat production was called *specific dynamic action* (p. 169) by Rubner. It is apparently due to reactions which the amino-acids undergo in the liver, or to the stimulation of other reactions in the cells by products formed in the liver. The phenomenon of specific dynamic action is not observed when the liver is removed. It is elicited by proteins and by individual amino-acids, but to a much smaller extent by corresponding amounts of urea and glucose.

It must be emphasized that amino-acids are constantly being liberated into the blood from tissue proteins which are being broken down during normal metabolism. Such amino-acids are dealt with in the liver in exactly the same way as are amino-acids derived from the food, and their nitrogen is excreted as urea. The amount of urea derived from this source is, of course, much greater when tissue breakdown is proceeding



A B

C D

Fig. 22, 2

A. Section of liver tissue from a well-fed rat. Stained hæmatoxylin eosin. × 450.

B. Section of liver tissue from a rat fasted for 48 hours. Stained hæmatoxylin eosin. ×450.

C. Section of liver tissue from a well-fed rat photographed in the ultraviolet microscope. Unstained. ×1100.

D. Section of liver tissue from a rat fasted for 48 hours, photographed in the ultraviolet microscope. Unstained. ×1100.

Note the shrinkage in the size of the cells in B and D owing to loss of protein and other constituents of the cell cytoplasm.

[*Facing page* 336

rapidly, as for example in fever and other states of increased metabolism, in chronic wasting diseases and in starvation.

A completely new outlook on the metabolism of amino-acids and proteins has been made possible by the pioneer work of Schoenheimer and his colleagues in America who used amino-acids containing the nitrogen isotope ^{15}N.

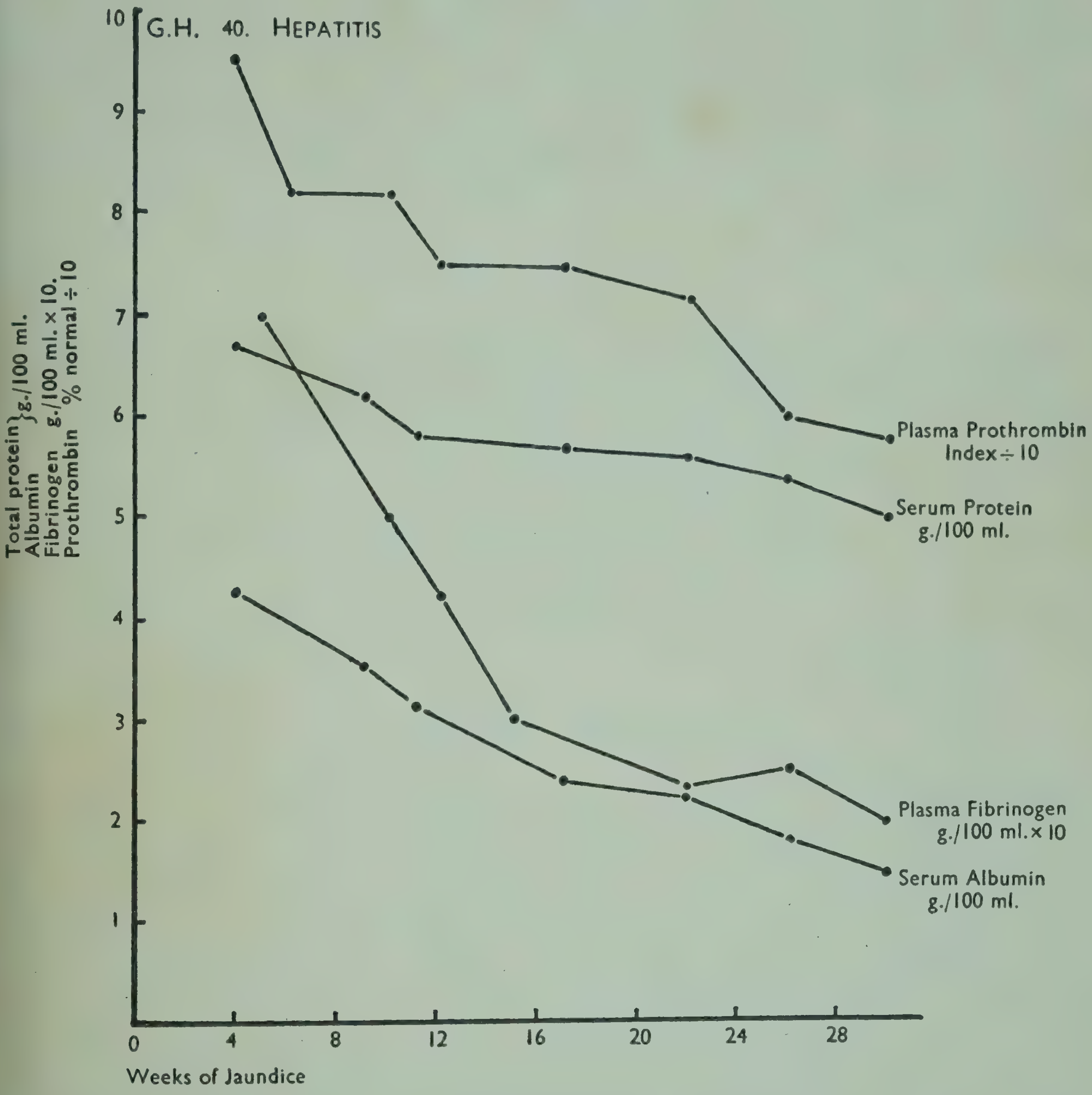


FIG. 22, 3

Changes in the concentration of various proteins in the blood of a male patient aged 40 with chronic disease of the liver and jaundice.

Such labelled amino-acids were fed to rats in nitrogen equilibrium (p. 188) on a diet containing adequate amounts of protein, and the excreta and the body tissues were subsequently analysed. According to classical theory most of the ^{15}N should have appeared directly in the urine as labelled urea, but this was not found to be the case. On the contrary most of the ^{15}N was traced to the tissue proteins of all organs examined, including even the skin (Table 22, 4).

When the individual amino-acids in these tissue proteins were isolated it was found that the ^{15}N was present in most of the amino-acids examined, except lysine. For example, when labelled leucine was fed, ^{15}N was subsequently found not only in the leucine of the tissue proteins but also in the glycine, tyrosine and arginine, and in particularly large amounts in the dicarboxylic amino-acids, glutamic acid and aspartic acid, which are active in transamination. Lysine never contained any labelled N. It appears, therefore, that the nitrogen of one amino-acid can be transferred to another amino-acid and that the amino-acids of the tissue proteins can be quickly replaced by their dietary analogues. This process involves the rapid breaking and reforming of peptide linkages. The peptide bonds of the tissue proteins must therefore rapidly open and close as amino-acids are taken out of the protein molecule and new amino-acids of the same kind are inserted. This happens even when amino-acids are abundantly supplied in the diet and when there is no suggestion of nutritional amino-acid deficiency.

TABLE 22, 4

Relative amounts of ^{15}N *in protein nitrogen of various organs of the rat after oral administration of either* L-*leucine or glycine containing* ^{15}N *in the amino group. The values are given as a percentage of* ^{15}N *in the compound administered.* (*From R. Schoenheimer* (1946), *The Dynamic State of Body Constituents.*)

Tissue	L-leucine fed	Glycine fed
Blood serum . . .	1·67	1·78
Hæmoglobin . . .	0·29	0.46
Liver	0·94	1·40
Intestinal wall . .	1·49	0·98
Kidney	1·38	. .
Heart	0·89	. .
Spleen	1·10	. .
Testis	0·77	. .
Skin	0·18	. .
Muscle	0·31	0·29

Since lysine does not take up nitrogen from any other amino-acids to any appreciable extent, Whipple and his colleagues have made use of it to label plasma proteins and follow their metabolism. They fed dogs with lysine containing ^{15}N in the ϵ-amino-position together with an adequate supply of other essential but unlabelled amino-acids. The ^{15}N in the lysine of the plasma proteins eventually reached a sufficiently high level for it to be measured accurately even after considerable dilution. The labelled plasma proteins were biologically indistinguishable from proteins of normal isotope concentration. When they were injected, as plasma, into normal dogs, they passed out of the blood stream at an initially rapid but ever decreasing rate. This outflow was balanced by a simultaneous inflow of plasma protein from the tissues. In 24 hours 50 per cent. and in 6 days 75 per cent. of the labelled protein had passed out of the blood stream. The plasma proteins are therefore always in rapid exchange with a mobile pool of body protein, and, indeed, dogs on a protein free diet can be kept in good health in nitrogen equilibrium for many months if whole dog plasma is given intravenously at regular intervals.

In more recent experiments in which plasma proteins labelled with radioactive carbon (^{14}C) in the lysine residue were administered intravenously to dogs, labelled proteins containing ^{14}C were subsequently found in all the tissues. The transformation of plasma protein to tissue protein was accompanied by such a small loss of the isotope in the urine and the expired air that Whipple has concluded that complete breakdown of protein to the constituent amino-acids is unnecessary for the conversion; a less drastic cleavage to fragments of polypeptide size, which are subsequently reassembled, is considered probable.

The tissue proteins are therefore not in a static condition but are in a constant state of flux. Even the relatively inert proteins, such as those of tendons and skin, take part slowly in this continual exchange of amino-acids.

The labile nature of the body protein is one of the most striking discoveries which have resulted from the use of isotopes. Sprinson and Rittenberg using glycine labelled with ^{15}N have calculated the half life of the chief tissue proteins, that is, the time taken for half of the tissue protein to be renewed. In man, the half life of the total body protein is 80 days, that of the liver and serum proteins being about 10 days and that of the proteins of the other tissues, mainly muscle, being 158 days.

For the rat the corresponding figures are, total protein 17 days, proteins of liver, plasma and internal organs 6 to 7 days, carcass proteins 21 days.

AMINO-ACID BREAKDOWN

Subsequent to their absorption or their release from tissue proteins, amino-acids may be used in the liver and tissues for the synthesis of tissue or plasma protein, or they may be broken down in the liver by the process known as *deamination*, their nitrogen appearing in the form of urea. Amino-acids may also lose their amino groups in the process of *transamination*, in which the amino group is transferred to a keto-acid (p. 341). A further use to which amino-acids are put in the body is in the synthesis of other nitrogenous compounds such as hormones, purines, pyrimidines and nitrogenous bases. Most of these are formed either from amino-acids themselves or from their breakdown products such as urea or ammonia. These processes are indicated in Table 22, 5.

TABLE 22, 5

Fate of ingested amino-acids after absorption

1. Temporary storage in liver and tissue.
2. Synthesis of
    - A. Tissue proteins.
    - B. Other nitrogenous bodies, including
        - hormones.
        - purines and pyrimidines.
        - nitrogenous bases (for example choline).
        - creatine.
3. Oxidative deamination or deamination by transamination with formation of keto-acid and
    - A. Oxidation of keto-acid via pyruvic acid (antiketogenic).
    - B. Oxidation of keto-acid via acetoacetic acid (ketogenic).
    - C. Synthesis of glycogen (gluconeogenesis).

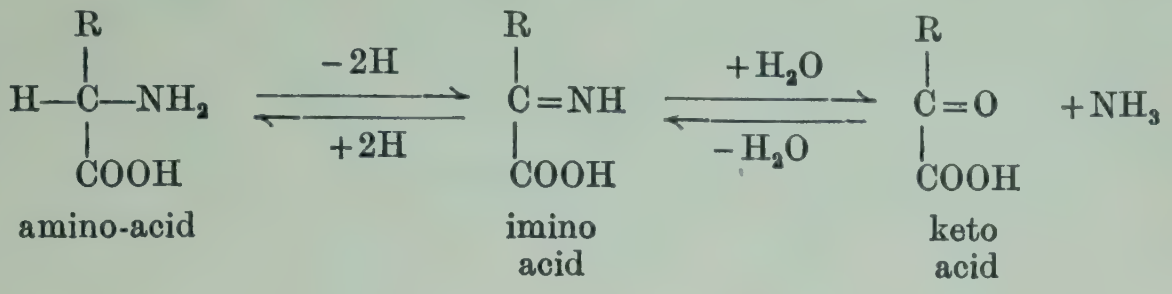
It should be understood that in the living body it is not possible to separate these processes one from another, since all are proceeding simultaneously so that sometimes one, sometimes another, is the more prominent according to the particular metabolic demands of the body at the time. Although the amino-acids, after absorption, may follow simultaneously several different pathways there are many factors such as age, rate of growth, energy requirements, protein content and caloric value of the diet, to mention only a few, which influence the course which their metabolism will take.

The urea which is formed from the amino group of the amino-acids cannot be metabolized further by the body and is ultimately excreted by the kidney. It will be clear from what has already been said that the urea in the urine is derived both from ingested protein and from tissue protein. It used to be held that the urea which came from the breakdown of amino-acids ingested in dietary protein was a product of what was termed *exogenous nitrogen metabolism.* On the other hand, a certain amount of urea was a product of what Folin termed *endogenous* or tissue metabolism. Thus the urea appearing in the urine when protein intake is reduced to zero is derived from the tissue proteins and is a product of *endogenous nitrogen metabolism.* This term covers all the metabolic processes involved in the building up and breaking down of the tissue proteins and other nitrogenous compounds in the body. In the normal individual on an everyday diet exogenous protein metabolism is always considerable, since more protein is ingested than is required to provide amino-acids for endogenous purposes.

The recent work with isotopes has given no support to this view that exogenous urinary nitrogen can be distinguished from endogenous urinary nitrogen. Schoenheimer, Ratner and Rittenberg stated in 1939 that "it is scarcely possible to reconcile our findings with any theory which requires a distinction between these two types of nitrogen. It has been shown that nitrogenous groupings of tissue proteins are constantly involved in chemical reactions; peptide linkages open, the amino-acids liberated mix with others of the same species of whatever source, diet or tissue. This mixture of amino-acid molecules, while in the free state, takes part in a variety of chemical reactions: some re-enter directly into vacant positions left open by the rupture of peptide linkages; others transfer their nitrogen to deaminated molecules to form new amino-acids. These in turn continuously enter the same chemical cycles which render the source of the nitrogen indistinguishable. Some body constituents like glutamic and aspartic acids and some proteins like those of liver, serum, and other organs are more actively involved than others in this general metabolic mixing process. The excreted nitrogen may be considered as a part of the metabolic pool originating from interaction of dietary nitrogen with the relatively large quantities of reactive tissue nitrogen."

DEAMINATION AND TRANSAMINATION

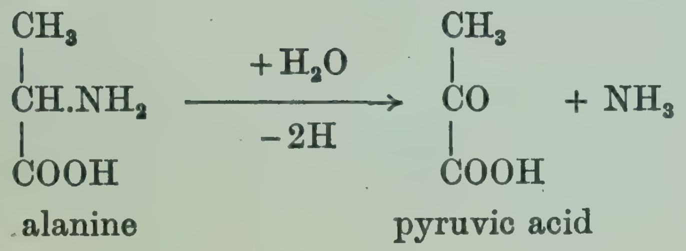
The first step in the breakdown of amino-acids in the liver is the removal of the amino group by the process of deamination and the formation of urea from the amino nitrogen. The non-nitrogenous residue of the molecule left after removal of the amino group is the corresponding keto-acid. The reaction probably proceeds by way of the corresponding imino-acid which hydrolyses spontaneously into the keto-acid and ammonia.



$$\underset{\text{amino-acid}}{\mathrm{H{-}\overset{\displaystyle R}{\underset{\displaystyle COOH}{C}}{-}NH_2}} \underset{+2H}{\overset{-2H}{\rightleftharpoons}} \underset{\text{imino acid}}{\mathrm{\overset{\displaystyle R}{\underset{\displaystyle COOH}{C}}{=}NH}} \underset{-H_2O}{\overset{+H_2O}{\rightleftharpoons}} \underset{\text{keto acid}}{\mathrm{\overset{\displaystyle R}{\underset{\displaystyle COOH}{C}}{=}O}} + \mathrm{NH_3}$$

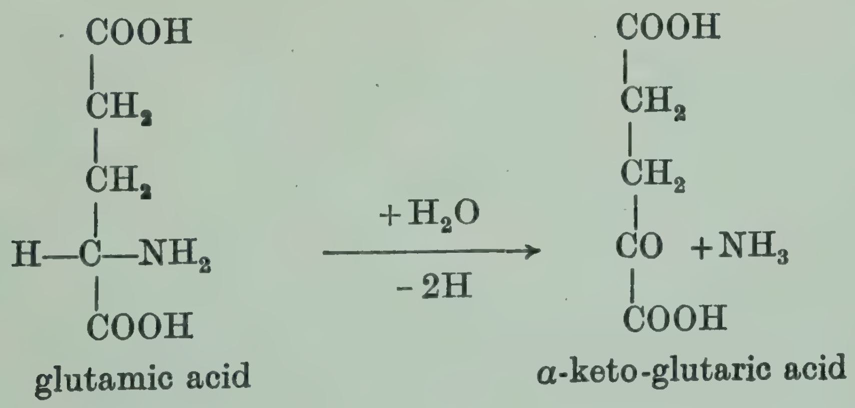
The reaction is catalysed by the L-*amino acid oxidase* (a flavoprotein) and since it involves oxidation by the removal of two atoms of hydrogen the process is usually referred to as *oxidative deamination*.

In the case of alanine, for example, the corresponding keto-acid is pyruvic acid:



$$\underset{\text{alanine}}{\mathrm{CH_3{-}CH.NH_2{-}COOH}} \xrightarrow[-2H]{+H_2O} \underset{\text{pyruvic acid}}{\mathrm{CH_3{-}CO{-}COOH}} + \mathrm{NH_3}$$

One of the most important deamination reactions is the removal of the amino group from glutamic acid. This reaction is catalysed by a specific enzyme, glutamic acid dehydrogenase which is particularly abundant in liver.

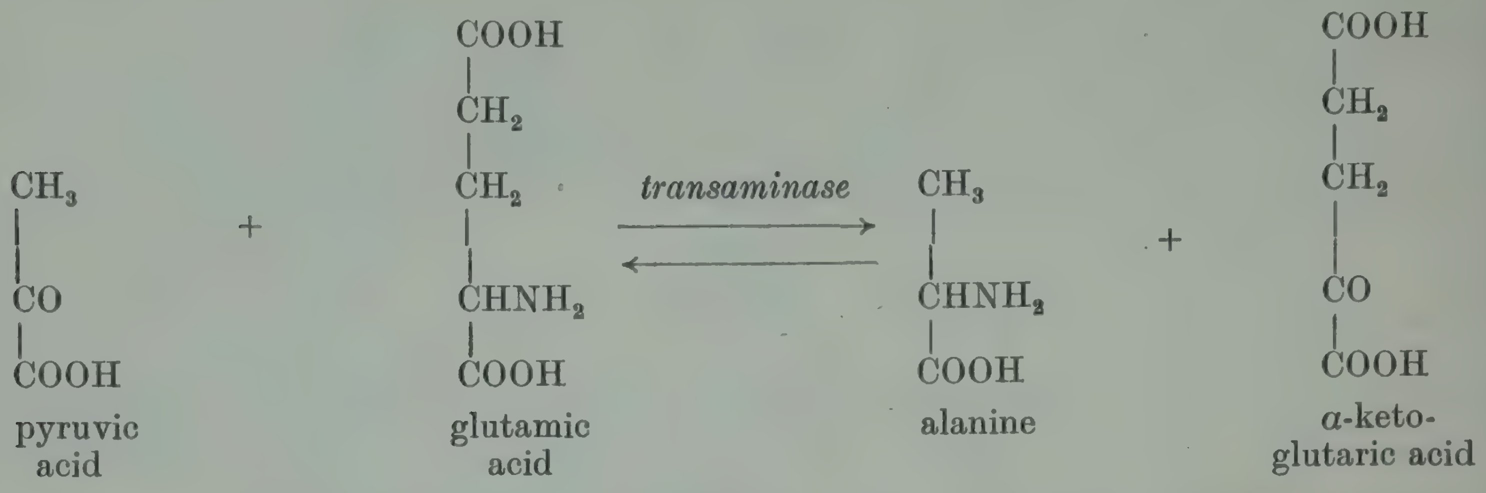


$$\underset{\text{glutamic acid}}{\mathrm{COOH{-}CH_2{-}CH_2{-}\overset{}{C}H(NH_2){-}COOH}} \xrightarrow[-2H]{+H_2O} \underset{\alpha\text{-keto-glutaric acid}}{\mathrm{COOH{-}CH_2{-}CH_2{-}CO{-}COOH}} + \mathrm{NH_3}$$

The process of deamination takes place chiefly in the liver but occurs also in the kidneys and in the intestinal mucosa. The ammonia is converted into urea; the keto-acid is oxidized to yield energy. These processes must be considered separately.

Transamination.—While the classical theory of amino-acid catabolism by deamination as described above has held sway for many years, recent work has made it clear that deamination is much less important than the process of *transamination* which must now be considered. In this process the amino group is transferred from an amino-acid to a keto-acid with the formation of a keto-acid corresponding to the original amino-acid and an amino-acid corresponding to the original keto-acid. This mechanism, therefore, plays a part both in the breakdown of amino-acids to yield keto-acids and in the formation of new amino-acids from keto-acids. It is catalysed by the *transaminase* enzymes of the liver with the aid of the coenzyme pyridoxal phosphate.

As an example of transamination we may take the reversible transfer of the amino group from glutamic acid to pyruvic acid to yield alanine and α-keto-glutaric acid.



In the process of transamination, in general one amino-acid must be glutamic acid and one keto-acid must be α-ketoglutaric acid.

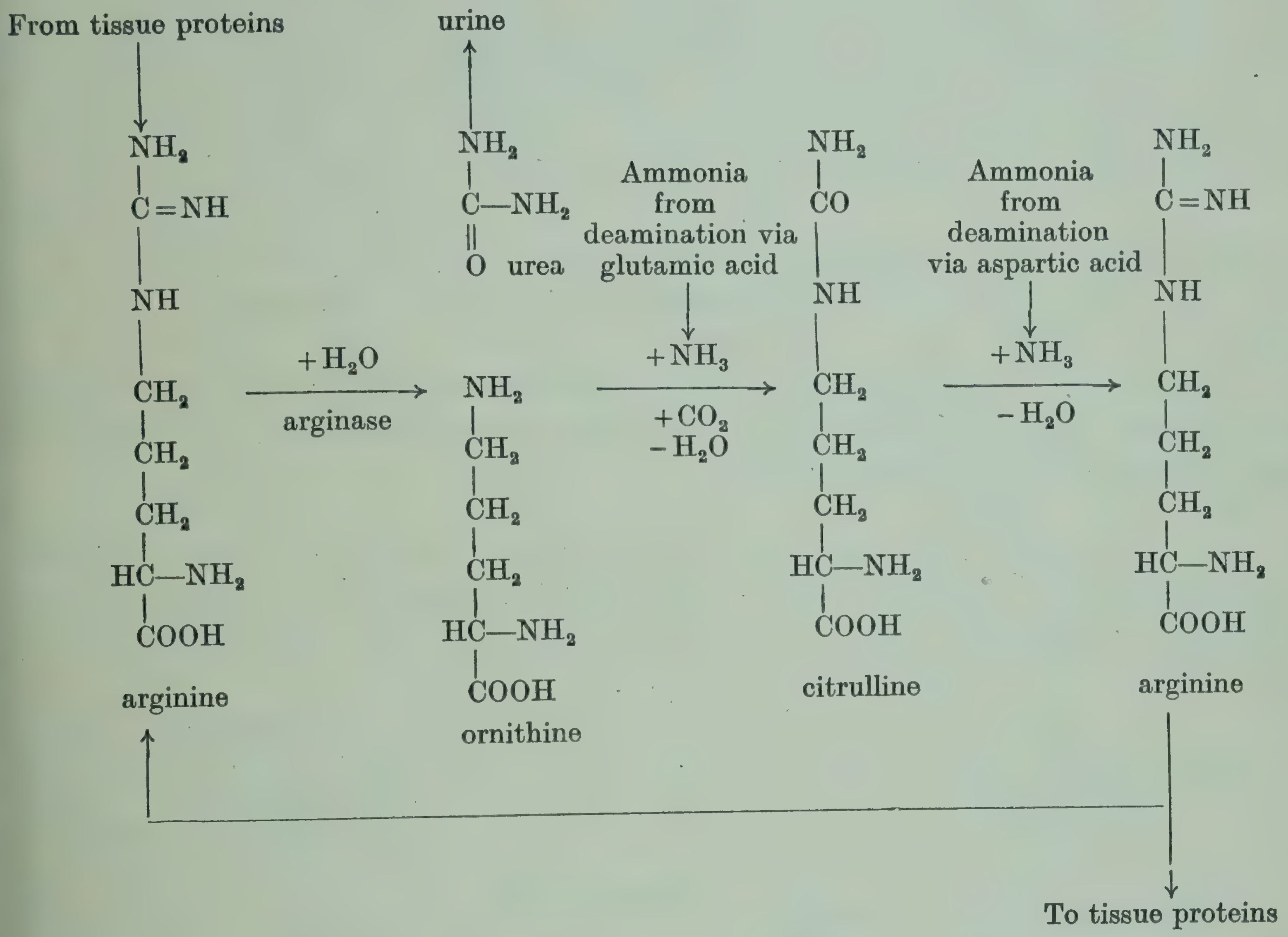
It should be noted that some of the keto-acids involved in these reactions, for example pyruvic acid and α-ketoglutaric acid, are involved in the processes of carbohydrate metabolism in the tricarboxylic acid cycle (p. 316). Their general disposal is discussed below.

UREA FORMATION

It has been conclusively proved by experiments in animals that the formation of urea occurs only in the liver. Bollman, Mann and Magath removed the livers from dogs and found that, during the short period for which the animals survived, amino-acids accumulated in the blood and urea excretion diminished. This was not due to failure on the part of the kidneys to excrete urea, since the blood urea level was also low. These conclusions were confirmed in experiments in which both liver and kidneys were removed. If the kidneys were removed first, the blood urea level rose; when the liver was extirpated the rise ceased, and the blood urea level remained constant so long as the animal survived. When both kidneys and liver were removed at the same time the blood urea remained at a constant level.

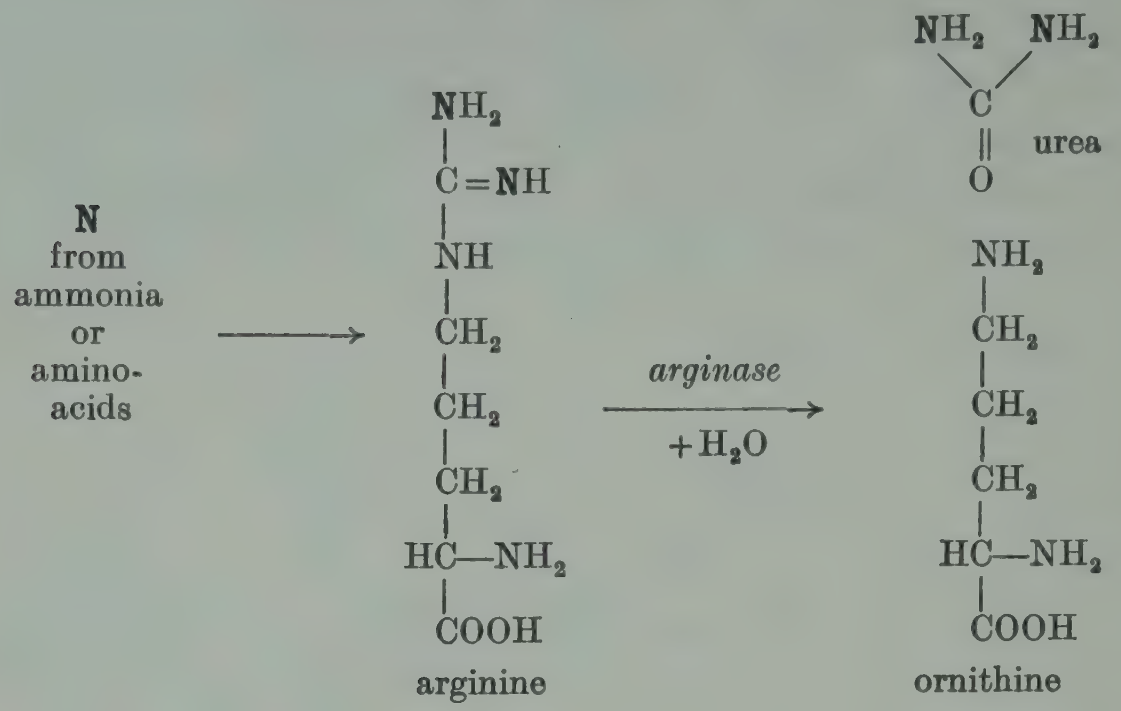
The function of the liver in urea production has also been demonstrated in experiments with the isolated organ or with liver slices. When the isolated liver is perfused with blood to which ammonium carbonate has been added urea is formed quantitatively from the ammonia. Of the many tissues which have been examined in the form of slices incubated in Locke solution, liver alone is able to bring about the conversion of ammonia to urea.

The conversion of ammonia to urea in the liver is not a simple combination of ammonia with carbon dioxide and water to form ammonium carbonate which is directly transformed to urea. The process by which this most probably occurs is that known as the ornithine-arginine cycle, first proposed by Krebs. According to this scheme, the amino-acid arginine is decomposed by the enzyme arginase to yield one molecule of urea and one of ornithine. Ammonia and carbon dioxide combine with the ornithine to form citrulline. The addition of a further molecule of ammonia results in the formation of arginine which takes part in the cycle again. Since the enzyme arginase does not act readily on arginine bound in proteins we must assume that molecules of arginine are continually being liberated in the free state from the proteins of the liver by the opening of peptide bonds. These arginine molecules are broken down by arginase to ornithine and urea. The ornithine molecule then reacts with ammonia and carbon dioxide to give a molecule of citrulline and then of arginine again. The arginine molecule may either be reintroduced at once into the protein or it may first go through the cycle of urea formation repeatedly.


From tissue proteins
urine
NH_2
C=NH
NH
CH_2
CH_2
CH_2
HC—NH_2
COOH
arginine
$+H_2O$
arginase
NH_2
C—NH_2
O urea
NH_2
CH_2
CH_2
CH_2
HC—NH_2
COOH
ornithine
Ammonia from deamination via glutamic acid
$+NH_3$
$+CO_2$
$-H_2O$
NH_2
CO
NH
CH_2
CH_2
CH_2
HC—NH_2
COOH
citrulline
Ammonia from deamination via aspartic acid
$+NH_3$
$-H_2O$
NH_2
C=NH
NH
CH_2
CH_2
CH_2
HC—NH_2
COOH
arginine
To tissue proteins

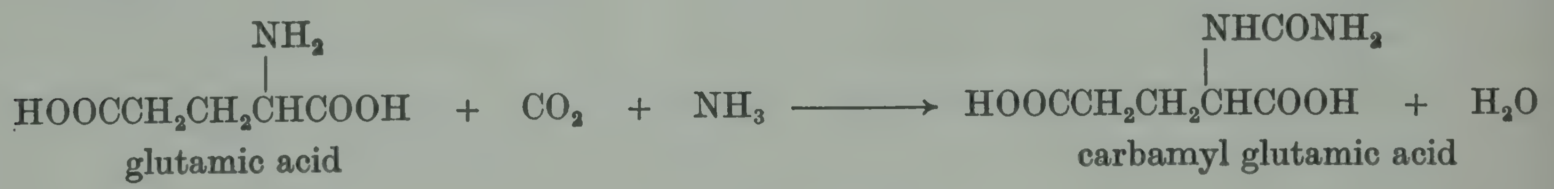
As evidence in favour of this mechanism Krebs pointed out that slices of mammalian liver, but of no other tissue, when kept in oxygenated Locke solution, could convert ammonia into urea, and that the process could be accelerated by the addition of ornithine or arginine. The enzyme arginase, which is essential in the scheme, is found abundantly only in mammalian liver. Birds excrete nitrogen not as urea but as uric acid; their livers contain no arginase. Tadpoles excrete ammonia while frogs excrete urea. Metamorphosis is accompanied by the appearance of arginase in the liver.

Evidence in favour of the ornithine-arginine cycle has also been obtained with the aid of isotopes. By the use of labelled carbon it has been conclusively proved that the carbon of urea is derived from carbon dioxide. When ammonia or amino-acids containing ^{15}N are fed to animals the labelled nitrogen is subsequently found in the urea of the urine and in the arginine of the tissue proteins. Of the four nitrogen atoms in arginine only the two in the amidine group $\left(—C\begin{smallmatrix}\diagup NH\\ \diagdown NH_2\end{smallmatrix}\right)$ are involved. The other two nitrogen atoms remain in the amino-acids of the cycle. Moreover when arginine with ^{15}N in the amidine group is fed to animals the labelled nitrogen very quickly appears in the urinary urea.



While urea formation by some such mechanism as this almost certainly does occur, the possibility that some urea is formed by other routes as well cannot at present be excluded.

Recent American work has indicated that the process of urea formation is more complicated than is indicated above. Glutamic acid reacts with carbon dioxide and ammonia to form carbamylglutamic acid, which in turn reacts with ornithine to form citrulline and glutamic acid.

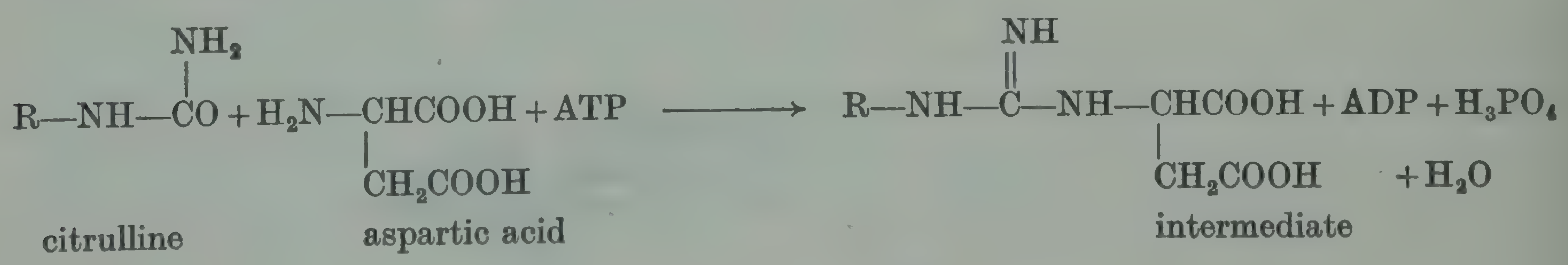


$$\underset{\text{glutamic acid}}{HOOCCH_2CH_2\overset{\overset{NH_2}{|}}{C}HCOOH} + CO_2 + NH_3 \longrightarrow \underset{\text{carbamyl glutamic acid}}{HOOCCH_2CH_2\overset{\overset{NHCONH_2}{|}}{C}HCOOH} + H_2O$$

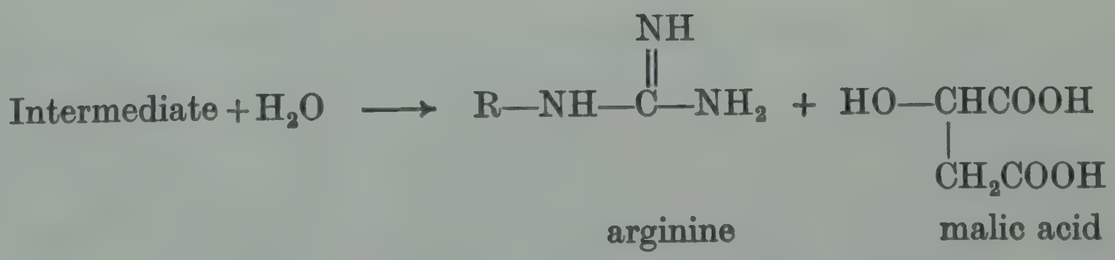
$$\text{carbamylglutamic acid} + \text{ornithine} \xrightarrow{\textit{transcarbamylase}} \text{glutamic acid} + \text{citrulline.}$$

The first of these reactions requires ATP, magnesium ions and aerobic conditions. The second reaction requires the enzyme *transcarbamylase.*

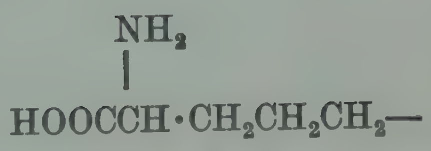
The citrulline then reacts with aspartic acid to form an intermediate compound which breaks down to arginine and malic acid, thus:—



$$\underset{\text{citrulline}}{R—NH—\overset{\overset{NH_2}{|}}{C}O} + \underset{\text{aspartic acid}}{H_2N—\underset{\underset{CH_2COOH}{|}}{C}HCOOH} + ATP \longrightarrow \underset{\text{intermediate}}{R—NH—\overset{\overset{NH}{\|}}{C}—NH—\underset{\underset{CH_2COOH}{|}}{C}HCOOH} + ADP + H_3PO_4 + H_2O$$



$$\text{Intermediate} + H_2O \longrightarrow \underset{\text{arginine}}{R—NH—\overset{\overset{NH}{\|}}{C}—NH_2} + \underset{\text{malic acid}}{HO—\underset{\underset{CH_2COOH}{|}}{C}HCOOH}$$



where R is $HOOC\overset{\overset{NH_2}{|}}{C}H\cdot CH_2CH_2CH_2—$

Each of these reactions is catalysed by a separate enzyme.

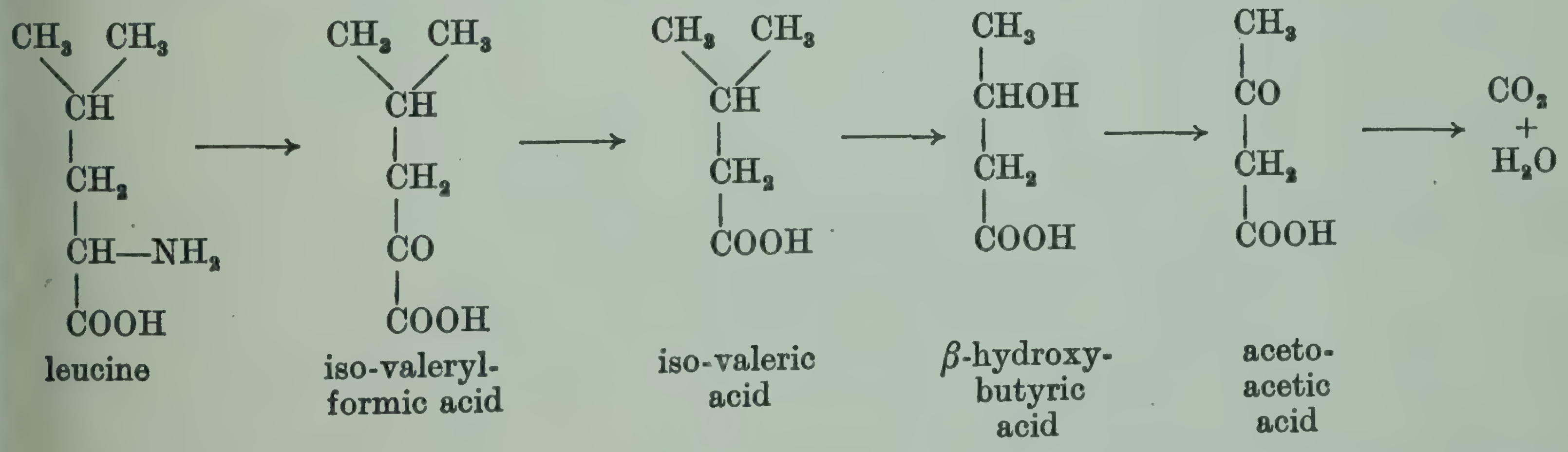
THE FATE OF KETO-ACIDS

As we have just seen, amino-acids yield keto-acids either by deamination or by transamination. The subsequent history of these keto-acids must now be considered. It should be emphasized that several of these keto-acids are involved in the tricarboxylic acid cycle as shown in Fig. 20, 19.

If amino-acids are fed one at a time to a starving phlorrhizinized dog it is found that some give rise to glucose in the urine, while others give rise to acetoacetic acid (p. 332). A few give rise to neither. In such an animal about 60 grams of glucose are formed and excreted in the urine for each 100 grams of protein metabolized. In other words, 60 per cent. of the protein is potentially glucogenetic. When the liver is removed, however, no such glucose formation takes place. Amino-acids can be divided therefore into three groups:

(1) Those of the first group, termed *antiketogenic* or *glycogenic* amino-acids, are deaminated to form keto-acids from which glycogen may be produced when necessary. They are oxidized to carbon dioxide and water in the same way as are carbohydrates. For example, alanine is deaminated to give the keto-acid, pyruvic acid, which is an intermediary product of carbohydrate breakdown. The antiketogenic amino-acids are glycine, alanine, serine, cysteine, cystine, aspartic acid, glutamic acid, valine, ornithine, arginine, proline and hydroxyproline.

(2) The second group consists of *ketogenic* amino-acids. After deamination they yield keto-acids which, during further oxidation to carbon dioxide and water, pass through the stage of acetoacetic acid. For example, leucine is deaminated to yield the keto-acid iso-valeryl-formic acid which is further broken down to the ketone bodies β-hydroxybutyric acid and acetoacetic acid.


CH_3 CH_3 CH CH_2 $CH—NH_2$ $COOH$
leucine
CH_3 CH_3 CH CH_2 CO $COOH$
iso-valeryl-formic acid
CH_3 CH_3 CH CH_2 $COOH$
iso-valeric acid
CH_3 $CHOH$ CH_2 $COOH$
β-hydroxy-butyric acid
CH_3 CO CH_2 $COOH$
aceto-acetic acid
CO_2 + H_2O

The ketogenic amino-acids are leucine, and possibly isoleucine, phenyl alanine and tyrosine.

(3) Amino-acids of the third group are neither ketogenic nor antiketogenic. They include methionine, tryptophan, lysine and histidine.

The fate of threonine, norvaline, norleucine and methionine is unknown.

The ketogenic or antiketogenic value of any particular protein depends of course on its amino-acid content.

AMMONIA FORMATION

While urea formation is, as we have seen, confined to the liver, the process of deamination can occur in the liver, in the kidney and in the intestinal mucosa. It also occurs as the result of bacterial action in the intestinal contents. The ammonia which is formed in the intestine and in the kidney passes in the blood stream to the liver for conversion to urea, but the amount of ammonia which is normally found in the blood is very small (0·1 to 0·2 mg. per 100 ml.). Unlike the blood, the urine contains appreciable amounts of ammonia and in conditions such as acidosis, where abnormally large amounts of acid are being excreted, the concentration of ammonium salts in the urine may be greatly increased. It is thought that the ammonia for the neutralization of excess acid is formed in the kidney itself, to some extent by deamination but mainly from deamidation of the amide glutamine. This view is supported by experiments with ^{15}N. When isotopic ammonia or isotopic urea is fed to animals the isotope can be recovered in the urinary urea but not in the urinary

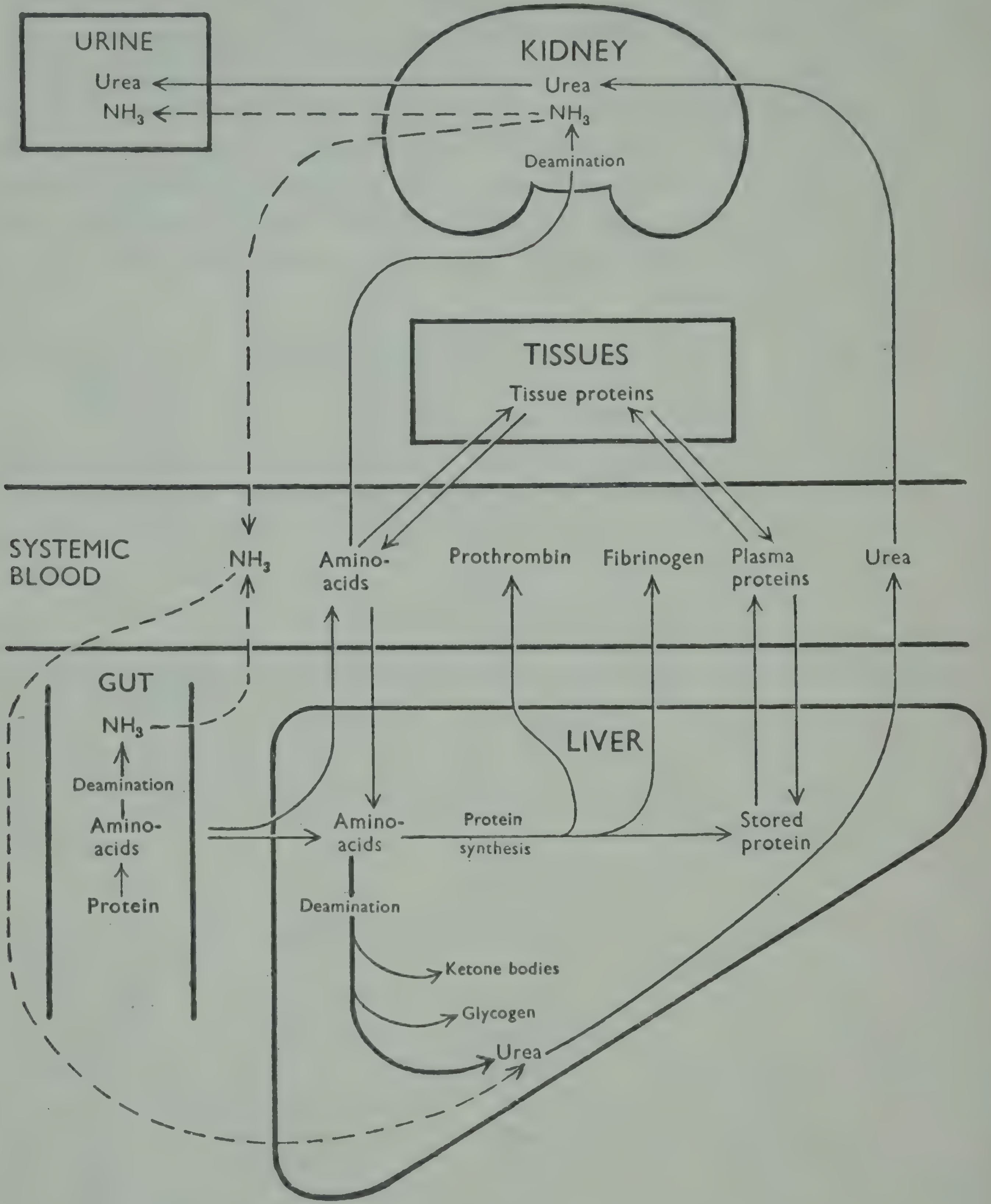

URINE
Urea
NH₃
KIDNEY
Urea
NH₃
Deamination
TISSUES
Tissue proteins
SYSTEMIC BLOOD
NH₃
Amino-acids
Prothrombin
Fibrinogen
Plasma proteins
Urea
GUT
NH₃
Deamination
Amino-acids
Protein
LIVER
Amino-acids
Protein synthesis
Stored protein
Deamination
Ketone bodies
Glycogen
Urea

Fig. 22, 6

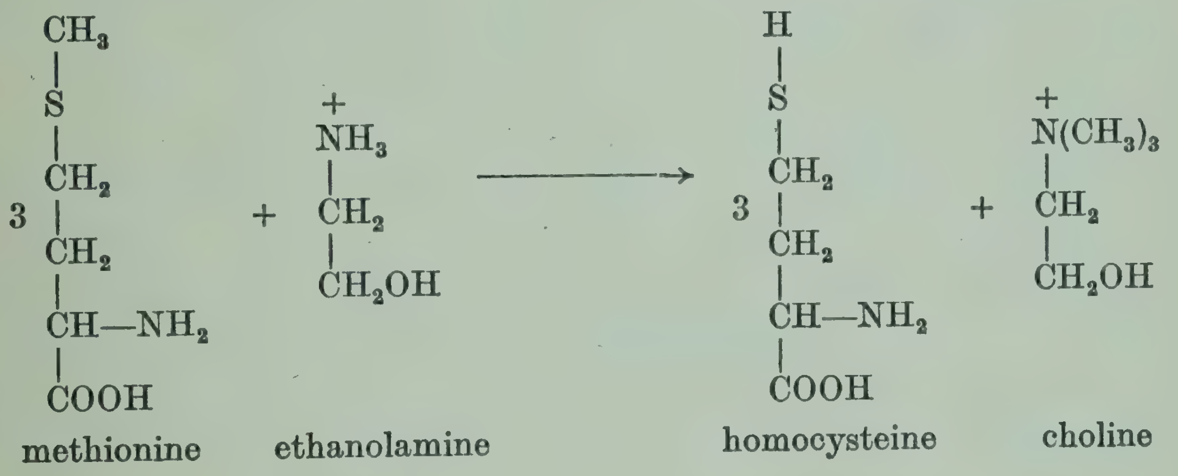
Diagrammatic representation of protein metabolism. Broken lines indicate pathways taken by ammonia.

ammonia. On the other hand, when certain isotopic amino-acids are fed, the isotope can be detected in the urinary ammonia.

The reactions involved in protein metabolism which have already been discussed are summarized in Fig. 22, 6.

OTHER REACTIONS OF AMINO-ACIDS IN THE BODY

1. **Transmethylation.**—In the important process, known as *transmethylation*, methyl groups may be transferred from the amino-acid methionine to other compounds, such as ethanolamine to form choline for phospholipid synthesis, or to guanidoacetic acid for the synthesis of creatine (p. 351). That such reactions do occur in the animal body has been proved by the use of methionine in which the hydrogen of the methyl groups was replaced by deuterium (heavy hydrogen). The deuterium was subsequently located in the methyl groups of the choline and of the creatine (see p. 352) of the tissues. The process of transmethylation is known to occur in man.



$$CH_3\text{—}S\text{—}CH_2\text{—}CH_2\text{—}CH(NH_2)\text{—}COOH + \overset{+}{N}H_3\text{—}CH_2\text{—}CH_2OH \longrightarrow H\text{—}S\text{—}CH_2\text{—}CH_2\text{—}CH(NH_2)\text{—}COOH + \overset{+}{N}(CH_3)_3\text{—}CH_2\text{—}CH_2OH$$

2. **Conjugation.**—Certain amino-acids serve as detoxicating agents by combining with a potentially harmful substance in the body to render it harmless and are excreted in this combined form in the urine. Such a process is known as *conjugation*. It involves the elimination of water and the appearance of the peptide linkage —CO.NH—. Thus, for example, benzoic acid and other aromatic acids absorbed from the alimentary tract or arising during the processes of metabolism are conjugated with glycine to form hippuric acid which is then excreted in the urine.

$$C_6H_5COOH + NH_2CH_2COOH \longrightarrow C_6H_5CO.NH.CH_2COOH + H_2O$$

benzoic acid — glycine — hippuric acid

This synthesis of hippuric acid takes place mainly in the liver and has been shown to occur less readily in the diseased organ. The amount of hippuric acid excreted after the intravenous injection of a standard dose of benzoic acid is therefore sometimes used clinically as an index of liver function (p. 419).

3. **Reactions of particular Amino-Acids.**—It will be remembered (p. 41) that many of the amino-acids are characterized by certain groupings or by cyclic nuclei of particular types. In the course of metabolism the amino-acids with aromatic nuclei give rise to compounds of great physiological importance, which will be discussed in greater detail elsewhere. As might be expected, amino-acids with these special configurations do not follow the metabolic pathways previously described for the simple

aliphatic amino-acids but are degraded by other routes which must now be briefly considered.

Two of the most important amino-acids are *tyrosine* and its related compound, *phenylalanine*, both of which are precursors of the hormones thyroxine and adrenaline. There is little doubt that tyrosine can be formed in the body from phenylalanine but the reverse process does not occur to any extent. On deamination these acids yield the corresponding keto-acids, *p*-hydroxyphenylpyruvic acid and phenyl pyruvic acid respectively (Fig. 22, 7). The latter can be further oxidized to 2 : 5-dihydroxyphenyl pyruvic acid. On shortening of the side chain this substance becomes homogentisic acid which after the ring is opened yields acetoacetic acid and malic acid. The latter is oxidized to carbon dioxide and water; the former yields two C_2-units which enter the citric acid cycle (Fig. 20, 19). Phenylalanine and tyrosine are therefore ketogenic acids.

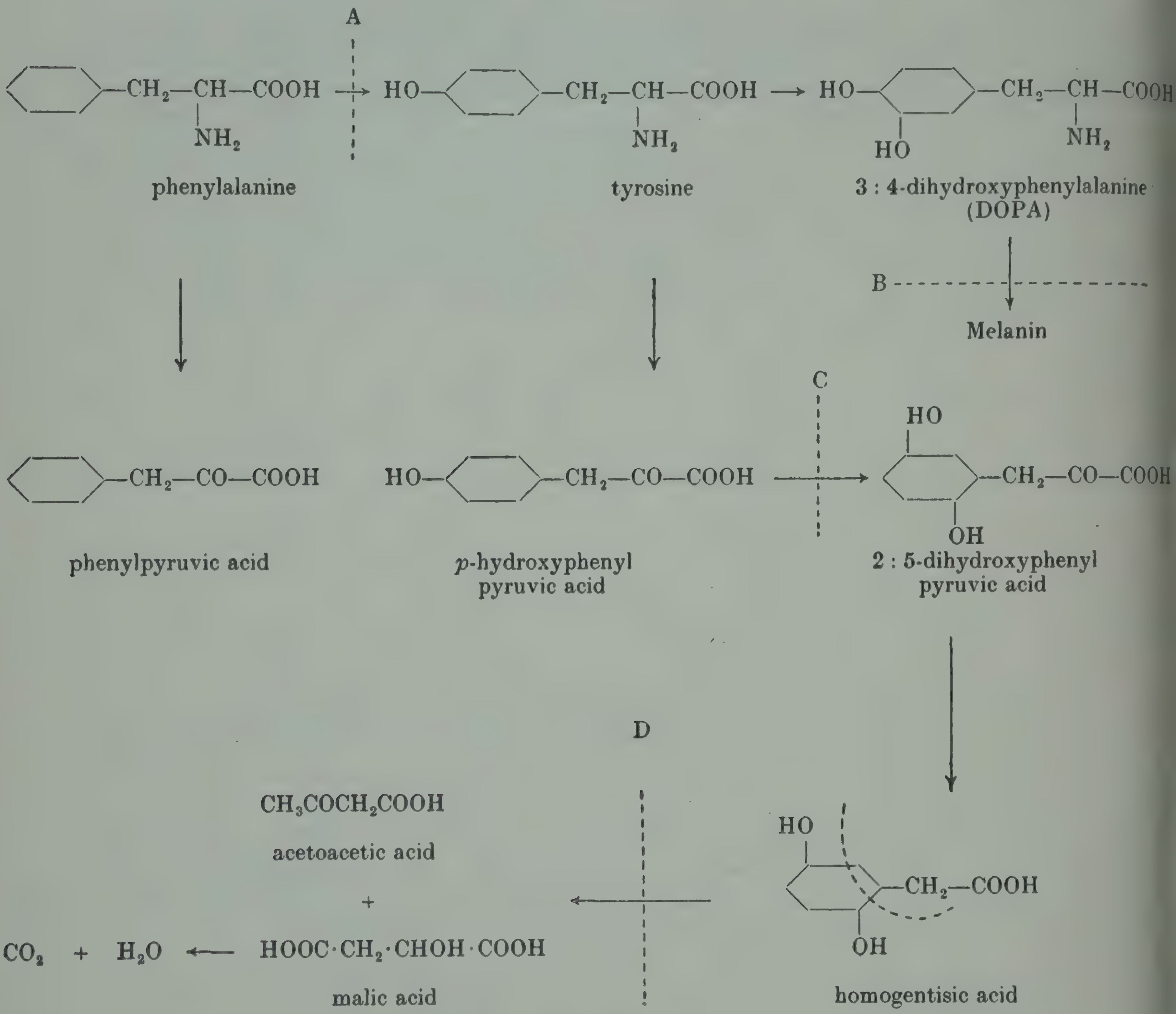

A
phenylalanine
tyrosine
3 : 4-dihydroxyphenylalanine (DOPA)
B
Melanin
C
phenylpyruvic acid
p-hydroxyphenyl pyruvic acid
2 : 5-dihydroxyphenyl pyruvic acid
D
CH₃COCH₂COOH
acetoacetic acid
+
CO₂ + H₂O
HOOC·CH₂·CHOH·COOH
malic acid
homogentisic acid

FIG. 22, 7

Metabolic pathways for phenylalanine and tyrosine. Blockage at A, due to absence of the appropriate gene and appropriate enzyme, results in phenylketonuria, at B in albinism, at C in tyrosinosis and at D in alkaptonuria. Blockage at C is found in ascorbic acid deficiency.

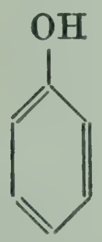
In the apparently harmless abnormality *alkaptonuria*, a rare inborn error of metabolism of these amino-acids, the process by which the ring of homogentisic acid is opened is defective. This acid therefore accumulates in the tissues and appears in the urine. Since the homogentisic acid, like many derivatives of tyrosine, is readily oxidized to black pigments of the melanin group, the urine of alkaptonuric persons slowly turns black when it is exposed to the air.

Tyrosine can also be oxidized by a complex series of changes to the black pigment *melanin*, which occurs in the hair, in the choroid coat of the eye, in the dark skin of the negro, and in the skin of white people after exposure to strong sunlight. The first step in this process is the oxidation of tyrosine by the enzyme tyrosinase to dihydroxyphenylalanine (DOPA). This compound undergoes further changes to form indole derivatives which polymerize to melanin (p. 82). In certain tumours the production of melanin by such a series of reactions may be very active and *melanuria*, the excretion of a black pigment in the urine, may actually occur.

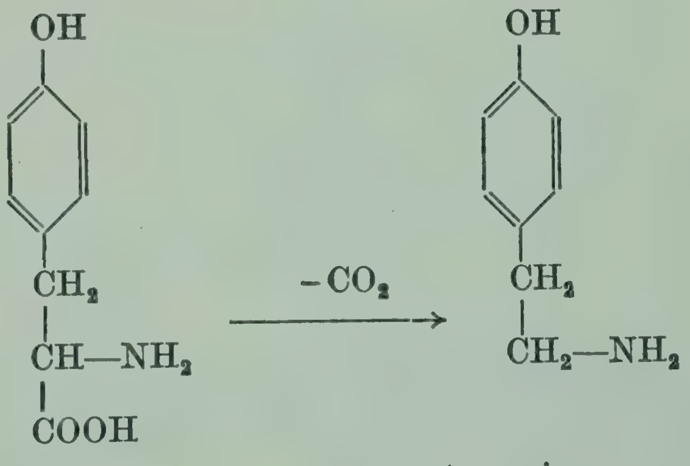
In a condition known as *phenylketonuria*, which occurs in certain mental defectives, the introduction of a hydroxyl group into the ring in phenylalanine does not take place but deamination to phenylpyruvic acid occurs. The latter compound cannot be further metabolized and appears in the urine, the amount being increased by administration of phenylalanine but not of tyrosine.

Further evidence of the metabolic pathway has been provided by the study of a patient with *tyrosinosis* in whom *p*-hydroxyphenylpyruvic acid could not be oxidized and appeared in the urine.

Under the influence of putrefactive bacteria in the intestine, tyrosine may be broken down to phenol; on decarboxylation, tyrosine yields the base tyramine.

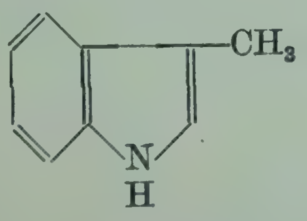


phenol

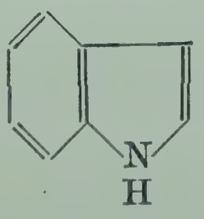


tyrosine tyramine

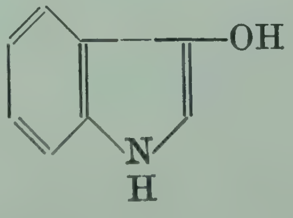
Tryptophan is probably the precursor of 5-hydroxytryptamine discussed in Chapter 51. Under certain conditions such as are found for example in the colon, it may be broken down to yield the evil-smelling compounds *skatole* and *indole*, and the related substance *indoxyl* which may appear in the urine as indican (see p. 355).



skatole



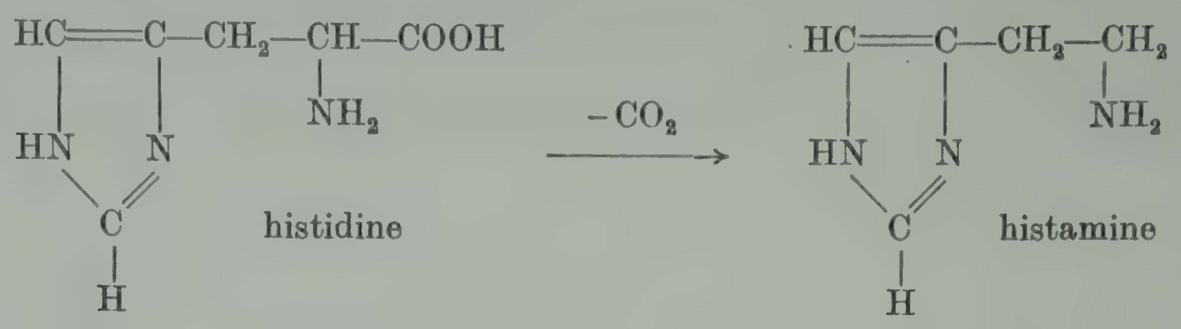
indole



indoxyl

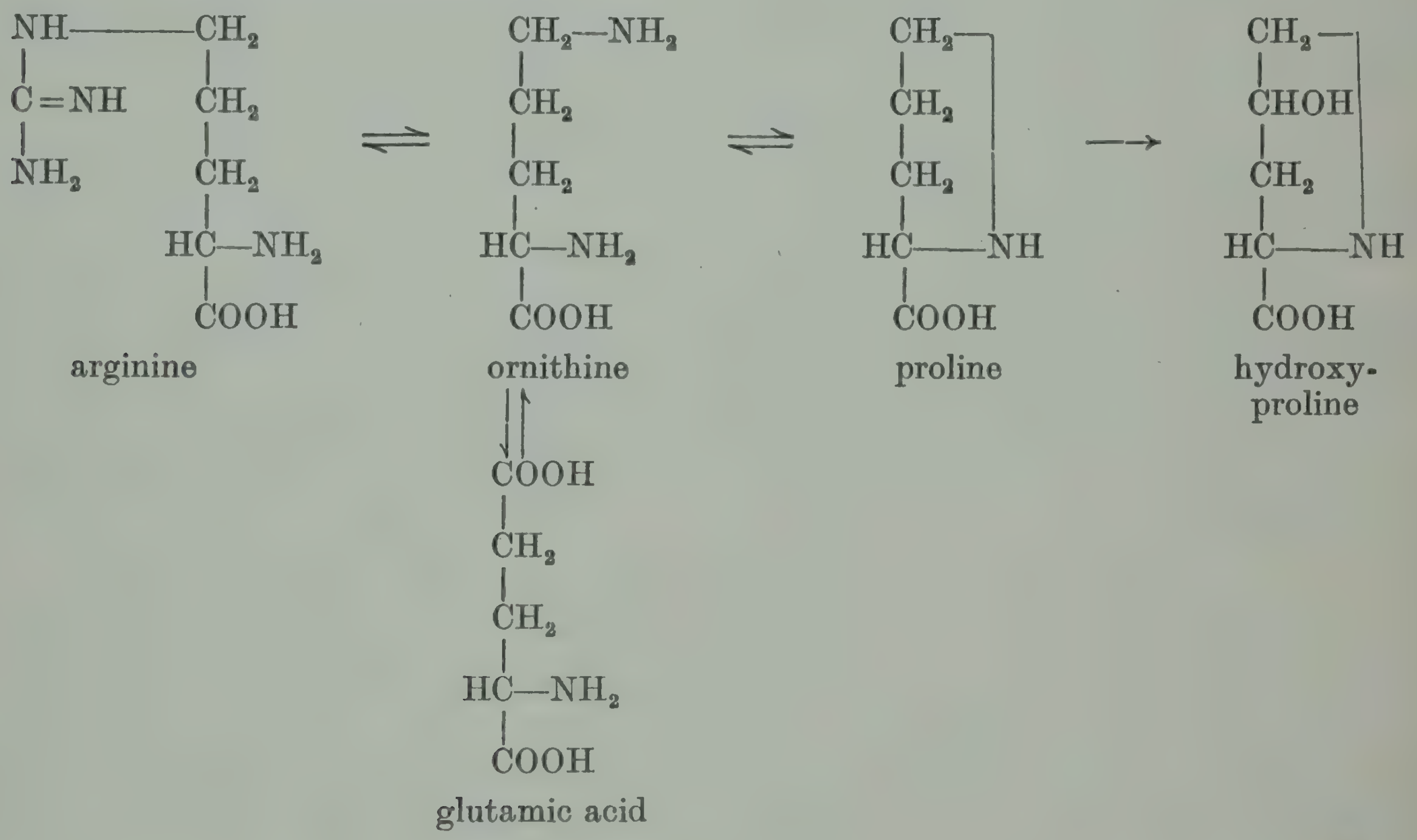
In certain species tryptophan gives rise to nicotinic acid.

The amino-acid *histidine* is the precursor of the physiologically important substance *histamine* which can be formed from it by the process of decarboxylation.

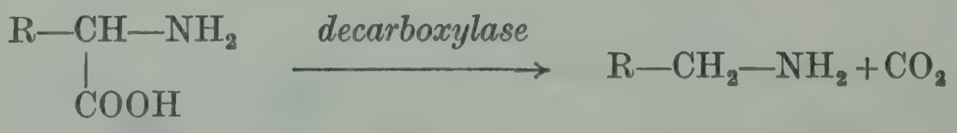

histidine
$-CO_2$
histamine

Apparently during pregnancy the metabolism of histidine is disturbed and unusually large amounts of the amino-acid appear in the urine. The concentration of histaminase in the blood is increased in late pregnancy.

By means of experiments with isotopes it can be shown that arginine, ornithine, proline and glutamic acid can be converted one into another, the carbon chain remaining intact.


arginine
ornithine
proline
hydroxy-proline
glutamic acid

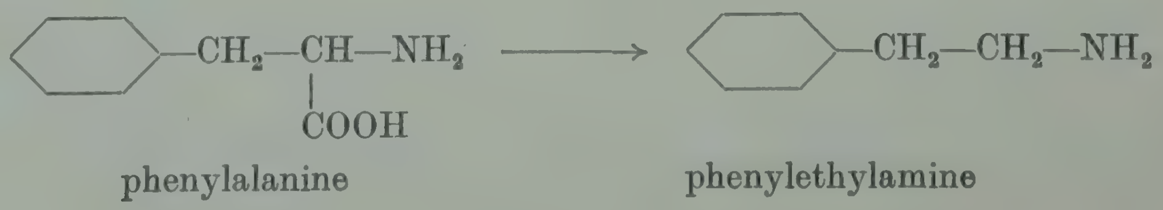
4. **Decarboxylation.**—Amino-acids are readily decarboxylated by the decarboxylases of certain bacteria to yield amines thus:



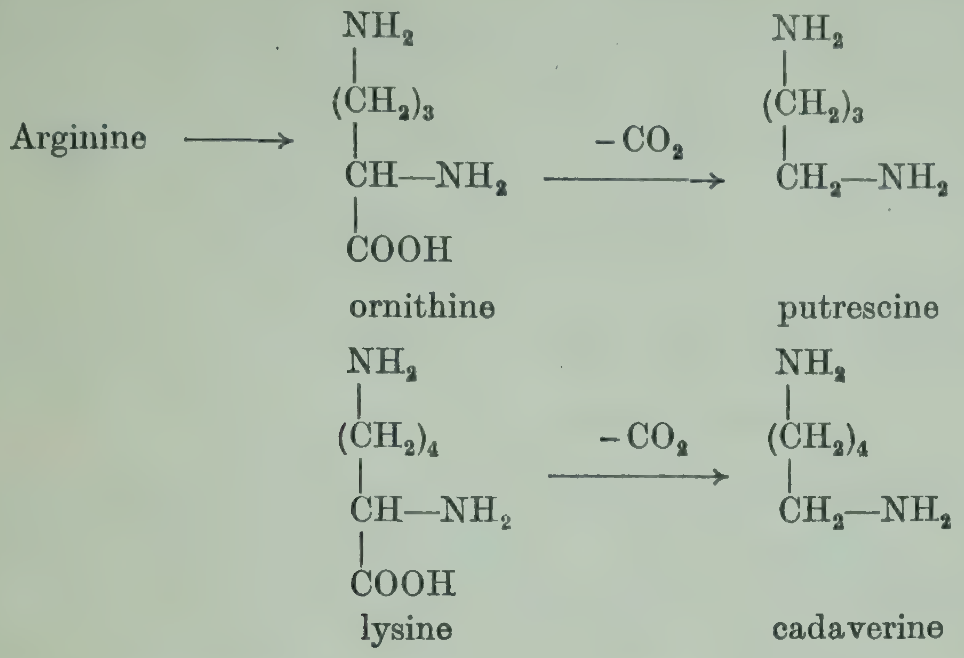
$$\underset{\textstyle COOH}{R—CH—NH_2} \xrightarrow{decarboxylase} R—CH_2—NH_2+CO_2$$

Pyridoxal phosphate (p. 129) acts as a co-enzyme in this reaction. Some amines formed in this way are of physiological importance.

We have already seen that *histidine* on decarboxylation yields *histamine* and *tyrosine* on decarboxylation yields *tyramine.* Phenylalanine on decarboxylation yields phenylethylamine.


phenylalanine
phenylethylamine

The diamino-acids *ornithine* (derived from arginine) and *lysine* may undergo decarboxylation to yield the corresponding diamines, *putrescine* and *cadaverine.*



$$\text{Arginine} \longrightarrow \begin{array}{l} NH_2 \\ | \\ (CH_2)_3 \\ | \\ CH{-}NH_2 \\ | \\ COOH \end{array} \xrightarrow{-CO_2} \begin{array}{l} NH_2 \\ | \\ (CH_2)_3 \\ | \\ CH_2{-}NH_2 \end{array}$$

ornithine putrescine

$$\begin{array}{l} NH_2 \\ | \\ (CH_2)_4 \\ | \\ CH{-}NH_2 \\ | \\ COOH \end{array} \xrightarrow{-CO_2} \begin{array}{l} NH_2 \\ | \\ (CH_2)_4 \\ | \\ CH_2{-}NH_2 \end{array}$$

lysine cadaverine

These amines are formed during the process of putrefaction in meat; they do not, however, produce 'food poisoning,' being relatively harmless to man.

ESSENTIAL AMINO-ACIDS

Nutritionally amino-acids are of two types, dispensable or indispensable. Indispensable or essential amino-acids are those which are "not synthesized by the animal organism out of the materials ordinarily available at a speed commensurate with normal growth" (Rose).

The indispensable amino-acids, both for man and for the rat, are valine, methionine, threonine, leucine, isoleucine, phenylalanine, tryptophan, lysine, arginine and histidine.

Some of these amino-acids, for example histidine, can be replaced in the diet by the corresponding keto-acid. In such a case the indispensable unit appears to be the carbon chain; the amino group can be introduced by the body as required. Indeed, all the indispensable amino-acids except lysine and threonine take part in the reversible transfer of amino groups from amino-acids to keto-acids.

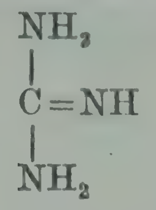
The work of Schoenheimer and others using isotopes has shown that the indispensable amino-acids fall into two classes. One class includes amino-acids, such as leucine and histidine, which continually yield and accept nitrogen. In such amino-acids the carbon chain is indispensable and cannot be synthesized by the animal but the amino group may be derived from some other molecule.

Indispensable amino-acids of the second class, lysine and threonine, form "a single indispensable chemical unit which has to be supplied as such to the animal" (Schoenheimer). Such a unit takes no part in the reversible transfer of nitrogen.

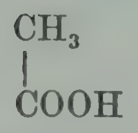
The importance of these indispensable amino-acids in connexion with the protein requirements in the diet is discussed on p. 185.

CREATINE AND CREATININE

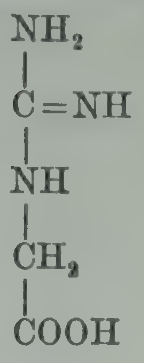
Two closely related compounds which must be discussed in relation to protein metabolism are *creatine* and *creatinine.* Creatine is methyl guanido-acetic acid; creatinine is its anhydride.



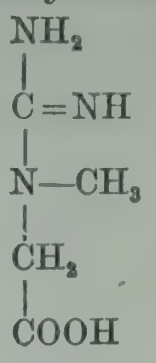
guanidine



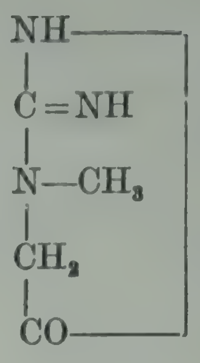
acetic acid



guanido-acetic acid (glycocyamine)



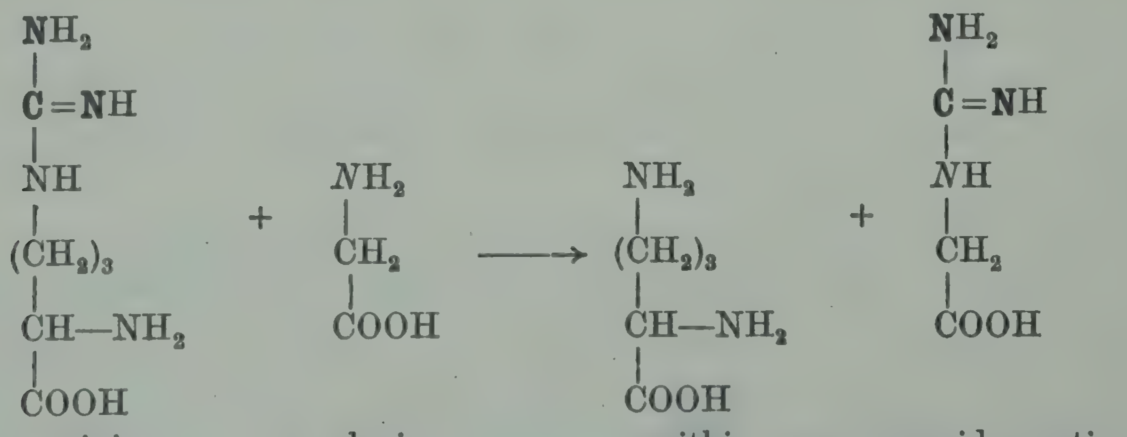
creatine



creatinine

Creatine is found abundantly in muscle but not in other tissues. Skeletal muscle contains about 0·5 per cent. creatine, mainly in the form of creatine phosphate or phosphagen, which is discussed in Chap. 42 in connexion with the chemical processes involved in muscular contraction.

Creatine is synthesized in the body and is not a necessary constituent of the diet. It is present in the muscles of herbivorous animals although none is contained in the vegetable tissues which they eat. With the aid of isotopes it has been shown that creatine is formed in the animal body from glycine, arginine and methionine. The glycine supplies the $=\mathrm{N}-\mathrm{CH_2}-\mathrm{COOH}$ portion of the molecule, arginine supplies the amidine group $(-\mathrm{C}\begin{matrix}\diagup\!\!\diagup\mathrm{NH}\\ \diagdown\mathrm{NH_2}\end{matrix})$ and methionine supplies the methyl group $(\mathrm{CH_3}-)$. When synthetic glycine containing the nitrogen isotope $^{15}\mathrm{N}$ is fed to rats, the labelled nitrogen appears in the appropriate position in the molecule of the muscle creatine. In a similar way it has been shown that arginine is the source of the other two nitrogen atoms in the creatine molecule. It is concluded that arginine and glycine react to form guanido-acetic acid thus:



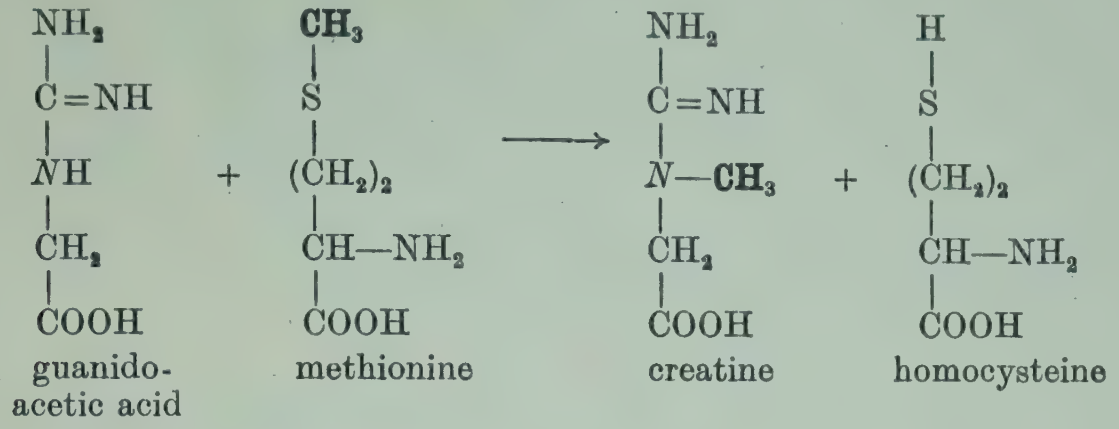
arginine glycine ornithine guanido-acetic acid (glycocyamine)

This transfer of the amidine group is known as *transamidination.*

Guanido-acetic acid is now known to occur in small amounts in both the tissues and the urine. After it has been fed, excretion of creatinine rises. The feeding of isotopic guanido-acetic acid results in an immediate formation of isotopic creatine and creatinine.

The methyl group of creatine is supplied by the amino-acid methionine by the process of transmethylation (p. 347). Creatine is rapidly formed when liver slices are incubated with guanido-acetic acid and methionine.

When rats are fed methionine containing heavy hydrogen (deuterium) in the methyl group, the heavy hydrogen appears subsequently in the methyl group of creatine isolated from the muscles. Choline may replace methionine as a source of the methyl group.



Creatine formation therefore occurs in two stages, the formation of guanido-acetic acid and its methylation, and there is evidence that the first stage occurs in the kidney and the second in the liver.

Some of the creatine phosphate in muscle loses water to yield phosphoric acid and the anhydride *creatinine* which appears in the urine. The use of labelled nitrogen has revealed that the urinary creatinine is formed from creatine (or its phosphate) and from no other source. Creatine, however, cannot be obtained from creatinine; the needs of the body for creatine are met by synthesis from methionine, arginine and glycine.

When creatinine is ingested most of it is rapidly eliminated in the urine, but when creatine is taken some is retained in the body, its fate being unknown. Some 20 to 30 per cent. of the ingested creatine is, however, excreted unchanged, only about 2 per cent. appearing in the urine as creatinine.

The urine of the normal healthy adult male contains creatinine and no creatine. The amount of creatinine excreted is about 1·0 to 1·5 grams per day and is remarkably constant for a given individual, and independent of the amount of protein in his diet. It is greater in muscular individuals, and appears to bear a relationship to the muscular development of the individual but not to his muscular activity. After severe exercise creatinine excretion may be increased but there is a corresponding fall in the excretion during recovery after exercise, so that the total daily output is unaffected.

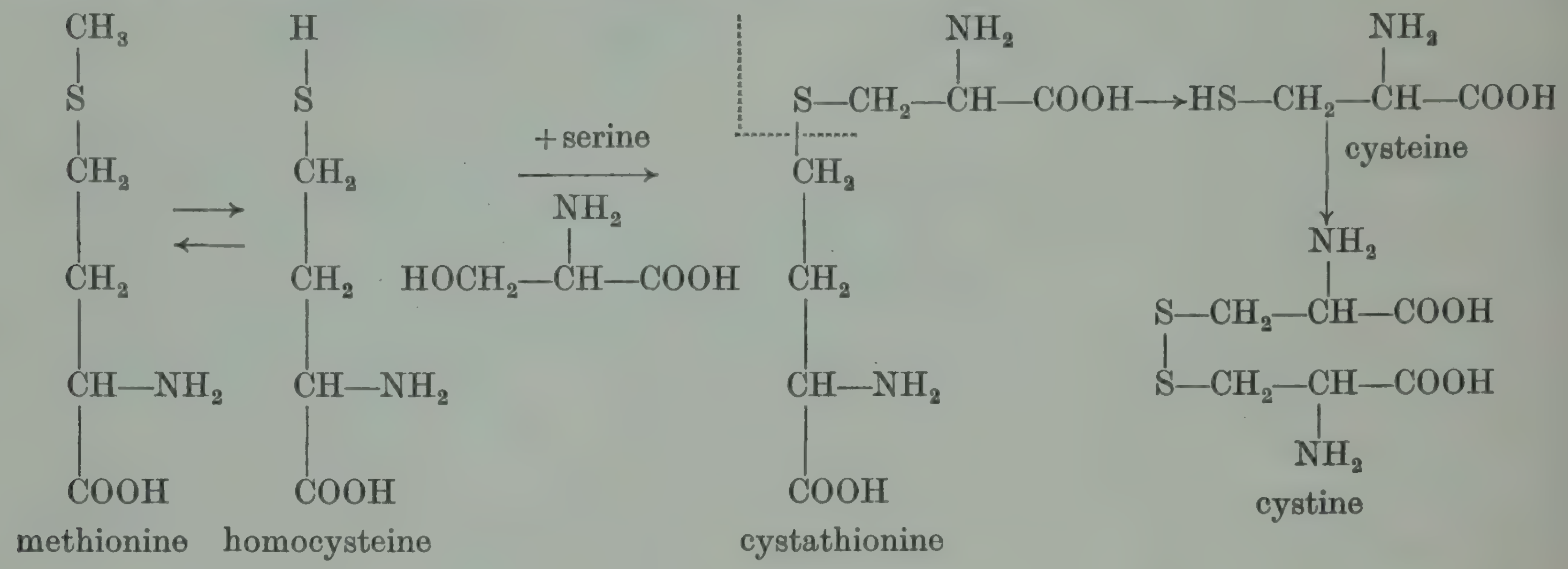
Creatine occurs normally in the urine of children along with creatinine. Creatine appears also in the urine of adults in whom breakdown of muscle tissue is occurring, that is in fevers, in prolonged starvation, and in hyperthyroidism. In women, creatine is excreted during menstruation and during and after pregnancy.

The urinary excretion of creatine following the ingestion of a standard dose (usually about 2 g. creatine hydrate) is determined in the *creatine tolerance test* which is used in the investigation of cases of myxœdema and hyperthyroidism (Chapter 51). In the former a smaller proportion (5 to 15 per cent.) than normal of the amount taken is excreted, while in the latter as much as 60 to 70 per cent. of the ingested creatine may be recovered from the urine.

In a group of rare metabolic disorders of muscles (including progressive muscular dystrophy and myotonia atrophica) creatinuria is a constant feature, the degree of the creatinuria running parallel with the patient's disability.

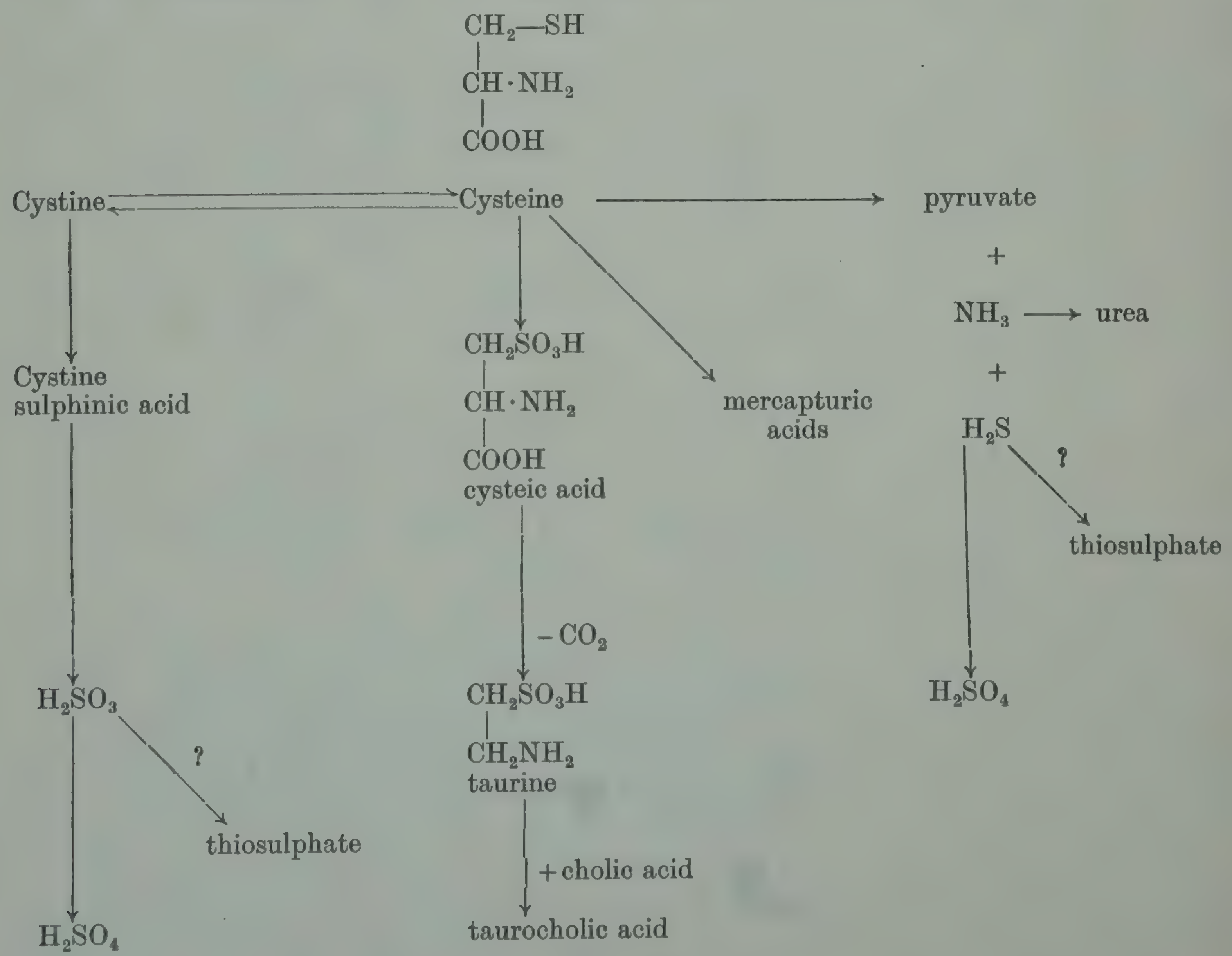
SULPHUR METABOLISM

The metabolism of sulphur compounds is conveniently considered along with that of the proteins. Sulphur is ingested chiefly in the form of the sulphur-containing amino-acids *cysteine*, *cystine* and *methionine*. Methionine may lose its methyl group in the process of *transmethylation* (p. 347), leaving *homocysteine*. This in its turn may be broken down to cystine by way of the intermediate compound cystathionine.



CH_3—S—CH_2—CH_2—CH—NH_2—COOH
methionine
H—S—CH_2—CH_2—CH—NH_2—COOH
homocysteine
+ serine
$HOCH_2$—CH(NH_2)—COOH
S—CH_2—CH(NH_2)—COOH
CH_2—CH_2—CH—NH_2—COOH
cystathionine
HS—CH_2—CH(NH_2)—COOH
cysteine
S—CH_2—CH(NH_2)—COOH
S—CH_2—CH(NH_2)—COOH
cystine

The fates of these acids are as follows: (1) Some of the ingested sulphur-containing amino-acids are used to repair or to build up the tissue proteins. Hair, for example, contains large amounts of cystine.



CH_2—SH
CH·NH_2
COOH
Cystine
Cysteine
pyruvate
+
NH_3 → urea
+
H_2S
?
thiosulphate
H_2SO_4
Cystine sulphinic acid
CH_2SO_3H
CH·NH_2
COOH
cysteic acid
mercapturic acids
– CO_2
CH_2SO_3H
CH_2NH_2
taurine
+ cholic acid
taurocholic acid
H_2SO_3
?
thiosulphate
H_2SO_4

Fig. 22, 8

The main pathways in sulphur metabolism.

(2) Small amounts of the sulphur-containing amino-acids are used in the synthesis of compounds of special significance, such as the hormone insulin, which is rich in sulphur-containing amino-acids, and the tripeptide glutathione (glutamyl-cysteyl-glycine), which plays a part in biological oxidations (see p. 89). Cysteine is utilized in the formation of the compound taurine, which combines with cholic acid (p. 27) to form the bile acid, taurocholic acid.

(3) Most of them, however, are broken down in the liver like other amino-acids. The nitrogen goes to form urea and the sulphur is oxidized to sulphuric acid, which appears in the urine as sulphate.

Little is known about the intermediate steps in the oxidation of the sulphur-containing amino-acids to sulphates, but some information has been derived from the study of a rare inborn error of metabolism, *cystinuria*, in which the patient constantly eliminates in the urine large quantities of cystine together with smaller amounts of other amino-acids. Although this condition is apparently compatible with perfect health the relative insolubility of cystine in urine not infrequently leads to its deposition in the urinary tract in the form of cystine calculi. A profound derangement of the metabolism of cystine is reflected in the disorder *cystinosis* in which deposits of this relatively insoluble amino-acid are found in many organs and tissues with profound effects upon growth and metabolism. The relationship between cystinosis, cystinuria and amino-aciduria (p. 357) is not at present clear.

Although information about the intermediate steps in the oxidation of the sulphur-containing amino-acids is incomplete, the sulphur compounds in the urine can be recognized as belonging to one of three types:—

(a) *Inorganic sulphate.*—This comprises much the greater part of the urinary sulphur.

(b) *Organic sulphate.* A small proportion of the urinary sulphur exists as sulphuric esters of certain phenolic bodies, for example, indoxyl formed in the process of detoxication (p. 420).

(c) *Neutral sulphur*, including for example thiosulphates, thiocyanates, cystine, mercapturic acids and urochrome.

NUCLEIC ACID AND NUCLEOPROTEIN METABOLISM

The digestion of the nucleoproteins and nucleic acids has already been discussed (p. 253). The end-products of nucleic acid digestion are the purine bases (or their nucleosides), the pyrimidine nucleosides, pentose and phosphoric acid. They are all absorbed into the blood stream as such and conveyed to the liver where they are further broken down.

The pentose is probably oxidized, but little is known about its fate.

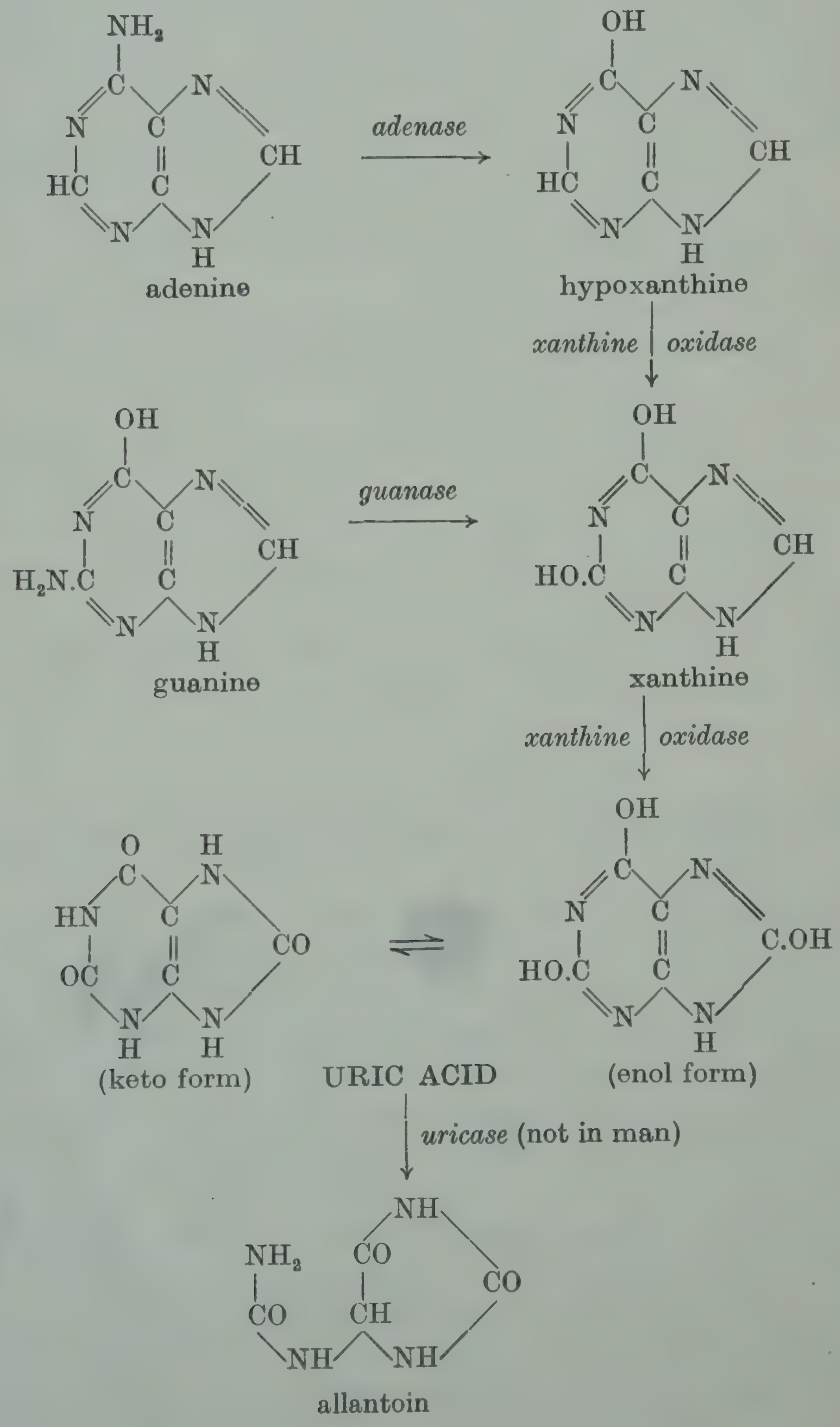
The pyrimidine bases are broken down to urea, ammonia, oxalic acid and carbon dioxide.

The purine bases, adenine and guanine, pass through a recognized series of changes resulting in the formation of *uric acid*. Adenine is deaminated by the enzyme adenase to yield hypoxanthine. Guanine in the same way is deaminated by guanase to give xanthine. Further oxidation is brought about by the enzyme xanthine oxidase in the liver. This enzyme oxidizes hypoxanthine to xanthine, and xanthine to uric acid, which is excreted by the kidney.

Uric acid therefore is the end-product of purine metabolism in man, and the amount excreted per day in the urine varies widely from 0·1 to 2·0 g.

Most mammals, however, do not excrete uric acid. They are able to oxidize it a stage further under the influence of the enzyme uricase of the liver to the compound *allantoin*, which therefore appears in the urine as the end-product of purine metabolism.

Man and the anthropoid apes are unable to oxidize uric acid in this way and, as a result, their urine contains little or no allantoin. The Dalmatian coach hound resembles other dogs in being able to oxidize uric acid to allantoin but, since its kidney is peculiar in showing no tubular reabsorption of uric acid, it excretes appreciable amounts of uric acid in the urine.


adenine
adenase
hypoxanthine
xanthine oxidase
guanine
guanase
xanthine
xanthine oxidase
(keto form)
URIC ACID
(enol form)
uricase (not in man)
allantoin

Uric acid is also excreted in large amounts by birds and reptiles, but in them it is the end-product not only of purine metabolism but of protein metabolism in general. Instead of excreting urea, they excrete uric acid as a semi-solid paste. They can thus conserve water by avoiding the necessity of producing urine in the form of a dilute aqueous solution.

When isotopic ammonium salts are fed to pigeons isotopic uric acid is excreted; when they are fed to rats isotopic urea is excreted. In pigeons the arginine of the proteins does not become labelled; in rats it is labelled. Arginine is therefore not involved in uric acid production in birds. When labelled urea is fed to pigeons there is little isotope in the uric acid subsequently excreted.

It will be obvious, from what has already been said, that a large proportion of the urinary uric acid in man comes from the purines ingested in the diet, which are especially abundant in such foods as liver, kidneys and sweetbreads. But even when the diet contains little or no purine, uric acid continues to appear in the urine. It is apparently derived mainly from the breakdown of tissue nucleoproteins, but a small amount may be synthesized in the body from urea and other sources.

The blood normally contains 2·0 to 3·5 mg. of uric acid per 100 ml. This value tends to be increased whenever cells are being destroyed and nucleoprotein liberated. Raised levels are therefore encountered in such conditions as leukæmia, polycythæmia, and in pneumonia during the process of resolution. In chronic renal disease the excretion of uric acid, like that of other nitrogenous bodies, is impaired and retention of uric acid occurs. Probably the highest values for uric acid in the plasma are found in gout immediately before an acute attack. In this disease salts of uric acid are actually deposited in the tissues, causing the typical chalky swellings (tophi) but the source of the abnormally large amounts of uric acid found in gout is not fully understood.

EXCRETION OF NITROGEN AND SULPHUR

From what has already been said it is clear that the output of nitrogen and sulphur in the urine is determined by the protein intake. Table 22, 9 shows the effect on a human subject of a change from a diet containing a moderate amount of protein to one low in protein. The result is a sharp reduction in the excretion of nitrogen affecting urea, ammonia and to a lesser extent uric acid. Creatinine excretion, however, remains unaltered. The changes in the excretion of sulphur are similar to those of urea.

The effects of starvation on the excretion of nitrogen have been discussed on page 188.

The urine normally contains small amounts of amino-acids (0·5 g. N per day) but in the rare condition of *amino-aciduria* the excretion is much increased. Such patients eliminate daily in the urine a complex mixture of some ten to twelve different amino-acids together with glucose and protein. As might be expected, the output of amino-acids in the urine may be much increased in patients with severe disease of the liver.

THE DYNAMIC STATE OF BODY CONSTITUENTS

Recent work with isotopes, particularly that of Schoenheimer, has given us a totally new conception of the chemical state of living matter. The tissues of the body are in a constant state of flux. Peptide and ester linkages open and the fragments so liberated merge with materials absorbed from the gut into what Schoenheimer has termed a ‘metabolic pool’ of components indistinguishable as to origin. In this pool amino-acids are deaminated and their nitrogen is transferred to other previously deaminated molecules. The new amino-acids thus formed then enter into the vacant spaces in the large protein molecules. Some of the nitrogen of the amino-acids finds its way to the amidine group of arginine and so to the excretory product urea, or in conjunction with glycine and methionine to creatine and creatinine.

Table 22, 9

Excretion of nitrogen and sulphur in the urine of a normal male student (*Glasgow University, October 1955*) *on a high and a low protein diet*

	Day	Estimated Intake in Diet in grams		Total Daily Output in Urine in grams					Nitrogen balance assuming fæcal N =10 per cent. of intake
		Protein	N	Urea N	Ammonia N	Creatinine N	Total N	Sulphate S	
High protein intake	2	176	28·1	18·0	0·99	0·65	20·7	1·66	+4·6
	3	198	31·6	23·7	1·15	0·63	26·0	1·44	+2·4
	4	193	30·9	22·7	1·18	0·64	25·9	1·52	+1·9
Low protein intake	2	71	11·3	11·1	0·79	0·64	13·4	1·42	−3·2
	3	59	9·4	8·5	0·53	0·63	10·2	0·60	−1·7
	4	39	6·2	6·0	0·38	0·65	7·7	0·66	−2·1

Schoenheimer has likened the living organism to a military regiment which is a unit with a well-defined and highly organized structure. Although its size varies within narrow limits only, the individuals who compose it are constantly changing as new men join up and others leave or die after varying lengths of service. The numbers of men entering and leaving are approximately equal but the individuals differ.

The large molecules in the body are continually being broken down and built up again under the influence of enzymes. The degradative changes are balanced by synthetic processes which must be coupled to energy-yielding reactions such as oxidation or dephosphorylation. When these energy-yielding reactions cease, as at death, the synthetic processes can no longer take place, but the degradative reactions continue so that the balance is upset and the living structure collapses.

REFERENCES

Bach, S. J. (1952). *The Metabolism of Protein Constituents.* Oxford University Press.

Borsook, H. (1950). Protein turnover. *Physiol. Rev.* **30,** 206–231.

Davidson, J. N. (1953). *The Biochemistry of the Nucleic Acids,* 2nd ed. London: Methuen.

Lerner, A. B., & Fitzpatrick, T. B. (1950). Biochemistry of melanin formation. *Physiol. Rev.,* **30,** 91–123.

Sahyun, M. (1948). *Proteins and Amino-Acids in Nutrition.* New York: Reinhold.

Schoenheimer, R. (1946). *The Dynamic State of Body Constituents.* Harvard University Press.

CHAPTER 23

THE BLOOD

THE blood in the vessels, although red to the naked eye, actually consists of a pale yellow fluid called *plasma* in which *formed elements*, the red corpuscles, the white (that is transparent) corpuscles and the platelets are suspended.

The blood subserves so many purposes that it is impracticable to catalogue them all. Its chief functions are the transport of oxygen and food materials to all the cells of the body and the removal of carbon dioxide and waste materials to the organs of excretion. Apart from small amounts in solution in the plasma oxygen is carried by hæmoglobin contained in the red cells while carbon dioxide is transported partly in the plasma and partly in the red cells. The white cells, though present to only 0·1 per cent. of the concentration of the red are none-the-less also essential for life, being concerned mainly with the defence of the body against micro-organisms. The blood platelets have an important role in the arrest of hæmorrhage and in the process of coagulation.

The plasma, as well as containing food materials in transit to the tissues and waste products on their way to be excreted, carries chemical messengers, or hormones, by which the activities of certain organs are regulated and co-ordinated. Normal values for the concentration of certain constituents of blood are given in Table 23, 1.

TABLE 23, 1

Concentration of Certain Constituents of the Blood of Normal Persons

The figures obtained in 80 per cent. of normal healthy individuals fall between the limits given in this table when the determinations are made by the methods described by King.

Whole blood	mg. per 100 ml.		m-equiv per litre	
	Lower limit	Upper limit	Lower limit	Upper limit
Cholesterol	140	215	—	—
Chloride (as Cl^-)	275	311	77	87
Creatinine	0·1	1·2	—	—
Glucose	68	96	—	—
Non-protein N	29	43	—	—
Urea	16	35	—	—
Uric acid	1·6	3·9	—	—
Serum				
Calcium	9·6	10·9	4·8	5·4
Potassium	15·1	19·6	3·9	5·0
Sodium	318	342	138	149
Chloride (as Cl^-)	360	378	101	106

(After E. J. King (1951), *Microanalysis in Medical Biochemistry*, 2nd ed. London: Churchill).

One of the chief constituents of the plasma is the plasma protein, amounting to some 6 or 7 per cent. and consisting of three main groups of proteins. The plasma albumins, about 4 per cent., exert an appreciable osmotic pressure and play an important part in controlling the movement of fliud across the capillary membrane as described on p. 636. The plasma globulins, accounting for some 2 per cent., are concerned in the antibody defence mechanism against infection. Fibrinogen and prothrombin both play an important role in blood coagulation which helps to prevent undue loss of blood from the vascular system after injury. Since the coagulation of blood, which occurs within a short time after it leaves the blood vessels, interferes with the examination of blood and its constituents, it is important to be able to control clotting. This aspect of the physiology of blood must, therefore, be considered first.

BLOOD COAGULATION

When blood is withdrawn from a vein in the arm, and emptied into a test-tube, clotting takes place in about five minutes at room temperature, complete coagulation being indicated by failure of the blood to move when the tube is tilted or inverted. Although the time of coagulation depends to some extent on the precise conditions under which the examination

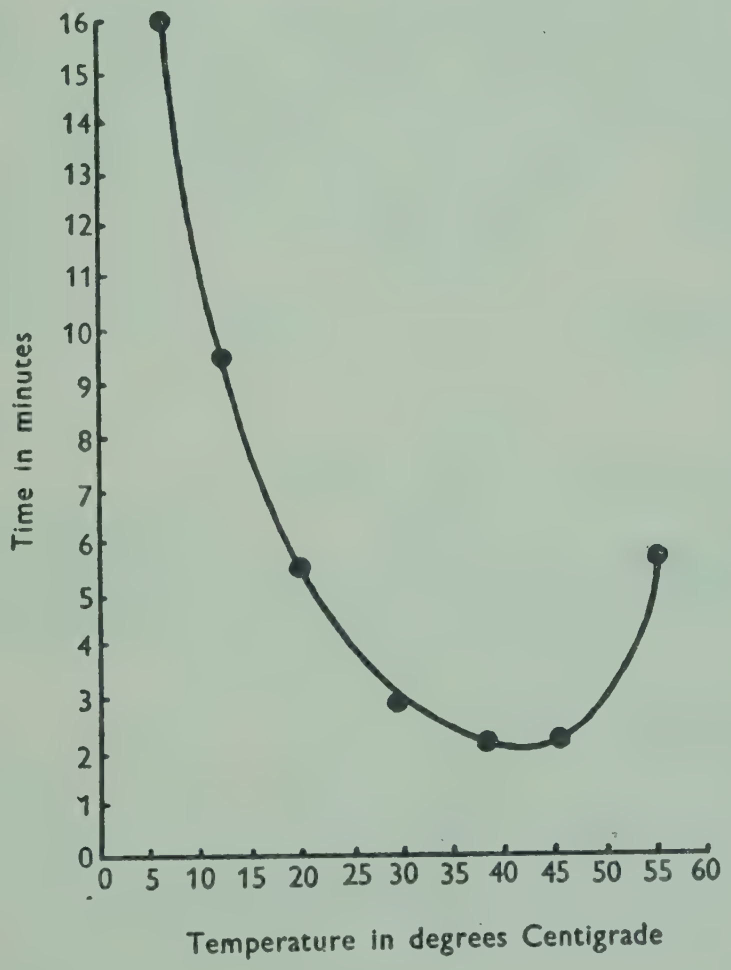


FIG. 23, 2

Curve showing the effect of temperature on the coagulation time of normal whole blood. Above 47° C clotting is poor since fibrinogen (the protein of the clot), which coagulates at 54° C, is partly precipitated before clotting can occur. Below 1° C the action of thromboplastin is much retarded and clotting is greatly delayed, so that the cells sink to the bottom of the container and the plasma can be removed. (R. G. Macfarlane (1948), *J. clin. Path.* **1**, 126).

is made, one of the main factors determining the rapidity of clotting is the temperature (Fig. 23, 2). The *coagulation* or *clotting time* may be measured in the following manner. One ml. of blood is transferred from a vein with a clean and perfectly dry syringe into a clean and dry test-tube 8 mm in diameter and about 7 cm long. The tube is stoppered and placed in a bath of water at 37° C from which it is removed and inverted every half minute until a firm clot appears. The time between taking the blood from the vein and clotting by this method is usually between three and five minutes but may extend to nine minutes. When blood clots in a test-tube a solid red mass is formed, but if the tube is left for some time the clot contracts, leaving a supernatant yellow fluid called *serum*. Microscopic examination of the clot shows it to be composed of irregularly arranged fibrils of a protein material called *fibrin*, in the interstices of which are found the formed elements; the red blood corpuscles entangled in the fibrin give the clot its red colour. If freshly shed blood is whipped with a glass rod or a fork, the fibrin as it forms adheres to the rod and the *defibrinated* blood (that is, the red cells and serum) remains fluid indefinitely. Chilling blood rapidly to 0° C prevents the disintegration of platelets and greatly retards coagulation.

Anticoagulants.—A number of other methods are now known for preventing the clotting of blood. Some of these are described below and an attempt has been made in Fig. 23, 3 to indicate the points at which the coagulation process is interrupted; the exact mechanism of action of the various anticoagulants cannot however be understood until the complex process of coagulation has been elucidated. The addition of sufficient sodium or ammonium oxalate, delays coagulation indefinitely, the red cells settle and the supernatant *oxalated plasma* can be pipetted off. Sodium

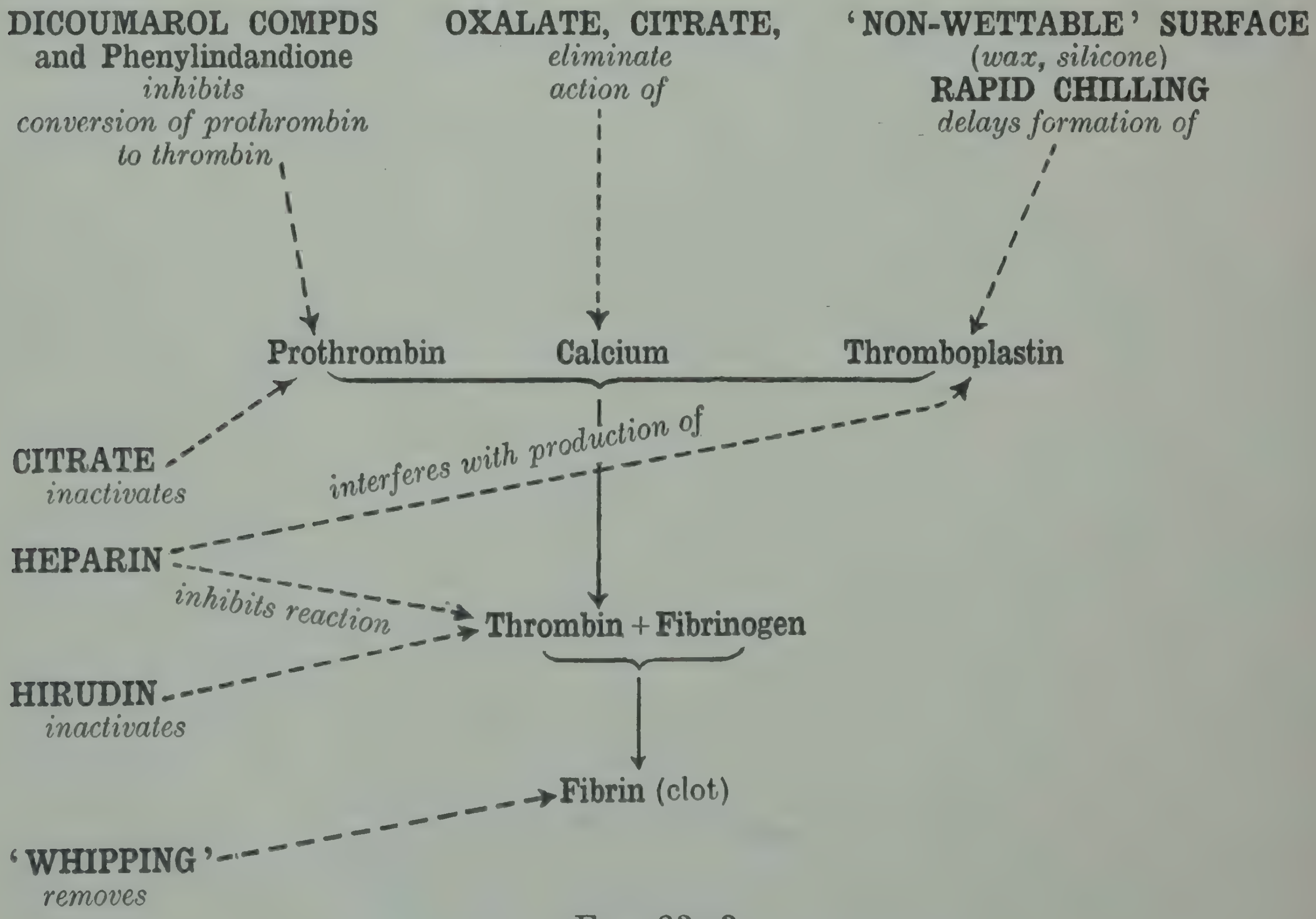

DICOUMAROL COMPDS and Phenylindandione
inhibits conversion of prothrombin to thrombin
OXALATE, CITRATE,
eliminate action of
'NON-WETTABLE' SURFACE
(wax, silicone)
RAPID CHILLING
delays formation of
Prothrombin
Calcium
Thromboplastin
CITRATE
inactivates
interferes with production of
HEPARIN
inhibits reaction
Thrombin + Fibrinogen
HIRUDIN
inactivates
Fibrin (clot)
'WHIPPING'
removes

FIG. 23, 3

Simplified scheme of the factors involved in the coagulation of the blood with an indication of the principal mode of action of some anticoagulants.

citrate, which has the great advantage over oxalate of being non-toxic, also prevents clotting. The inhibitory action of oxalate and citrate is complex and is not solely due to the removal of calcium ions from solution. Blood mixed with an equal volume of saturated sodium sulphate, a protein precipitant, is rendered incoagulable; the supernatant fluid which separates when the blood is allowed to stand is called *salted plasma.* For experimental purposes clotting can be prevented by various dyes, for example, chlorazol fast pink and by dextran sulphate of molecular weight 20,000 or less.

A second group of anticoagulants consists of those obtained from plants or animals. In 1916 McLean prepared an anticoagulant from liver which he called *heparin* (see p. 18). It has since been found in lung, muscle, heart and blood. The basophil granules of tissue mast cells and of the basophil polymorphs in the peripheral blood are probably composed of heparin, which is also believed to be produced in the mast cells of the intima of the blood vessels because they stain characteristically with toluidine blue. One mg. of pure heparin in 10 ml. of blood is sufficient to prevent clotting. Because of its low toxicity heparin is used clinically to prevent the formation of clots within the vessels in patients in whom such a tendency exists, but it disappears quickly from the blood stream, being partly excreted by the kidney and partly destroyed in the body. Its effect can be neutralized at once by injecting protamine (p. 53). Heparin interferes with coagulation at various points (Fig. 23, 3). It inhibits the inter-reaction of thrombin and fibrinogen and interferes with the formation of thromboplastin; it also potentiates the activity of an anticoagulant (antithrombin) normally present in blood. Both heparin and hirudin (see below) retard the agglutination and lysis of platelets.

TABLE 23, 4

Amounts of Anticoagulants used in various procedures to prevent Clotting of 10 ml. of Blood

Sodium citrate	60 mg.
Sodium oxalate	16 mg.
Sodium fluoride * Sodium oxalate	10 mg. 30 mg.
Potassium oxalate Ammonium oxalate	8 mg. 12 mg. †
Sodium citrate, 3 per cent. solution ‡	1 ml. to 9 ml. of blood.
Heparin	1 mg.

* Sodium fluoride is an enzyme poison; this anticoagulant mixture is used to delay for example, glycolysis in shed blood.

† With this mixture there is no alteration in the size of the cells due to the operation of osmotic factors; it is therefore used in determinations of cell volume.

‡ This solution of sodium citrate is isotonic with human red cells.

It was noted in 1929 that cattle feeding on spoiled sweet clover suffered from a hæmorrhagic disorder and in 1941 the toxic principle, *dicoumarol*, was identified and synthesized. This substance, which is quite inactive *in vitro*, interferes with the conversion of prothrombin by reducing the concentration of prothrombin and factor VII. Dicoumarol, and similar compounds, are extensively used in medicine to reduce the danger of intravascular coagulation, but, unlike that of heparin, their effects are not observed until some hours after administration. Bites made by a leech bleed freely and the animal is thus able to ingest large amounts of blood. The buccal glands of the leech contain the anticoagulant *hirudin* (an antithrombin). The venom of certain snakes contains a very powerful anticoagulant which acts by interfering with the action of thromboplastin.

If blood can be prevented from coming into contact with tissue fluid, or with substances which it can wet, clotting is greatly delayed. Blood collected in vessels coated with a thin layer of clean paraffin wax may remain fluid for half an hour. This phenomenon is made use of in the construction of tubes and cannulæ which, if made of perspex or nylon or coated with silicone, do not become blocked with blood clot so quickly as do glass cannulæ.

The actions of the commoner anticoagulants are summarized in Table 23, 3; some of them, however, act at more than one point in the complex process of coagulation which is, as yet, incompletely understood.

The Clotting Mechanism

Although the scheme given in Fig. 23, 3 was devised as long ago as 1904 by Morawitz to describe the basic facts, research work since 1940 has clearly indicated that the process is much more complex, and it is not at present possible to see how the various events are related. A satisfactory theory of blood coagulation must account for the fact that the blood remains fluid in the vessels. The smooth endothelium of the vessels does not encourage clotting and the platelets do not disintegrate to give thromboplastin. In addition it is likely that blood contains anticoagulants, an antithrombin and possibly heparin, which also prevents the action of thrombin.

From what has been said already it is clear that calcium is necessary for coagulation. If oxalated plasma has excess of calcium chloride added to it, clotting occurs. This experiment shows further that clotting occurs in the plasma and that the formed elements are not involved. The plasma clots owing to the appearance of *fibrin*. Fibrin cannot be present as such in the circulating blood but is produced by a physico-chemical change in a soluble precursor *fibrinogen* which is present normally in the blood plasma. This protein material, fibrinogen, can be precipitated from plasma by salting out or by alcohol precipitation. By repeated precipitation fairly pure fibrinogen can be obtained which has a molecular weight of 330,000. Since the fibrinogen content of the blood falls after hepatectomy it is usually assumed that fibrinogen is formed in the liver. A solution of fibrinogen cannot be made to clot by addition of calcium salts which must, therefore, act earlier in the sequence of events than at the stage fibrinogen → fibrin. Such a solution clots in a short time if serum alone is added; the substance responsible for the clotting of fibrinogen in the absence of calcium salts has been named *thrombin*.

This substance must act directly on fibrinogen since it causes clotting in oxalated plasma. Thrombin is almost certainly a proteolytic enzyme many of which can also form fibrin from fibrinogen, but it also has a powerful accelerating effect on coagulation because of its property of causing lysis of platelets.

Thrombin cannot, however, be present as such in circulating blood or in fresh oxalated plasma, because although both contain fibrinogen they do not clot spontaneously. Actually a precursor of thrombin, *prothrombin*, is present in circulating blood and the change from prothrombin to thrombin occurs in the presence of calcium and of *active plasma thromboplastin* (thrombokinase).

Active plasma thromboplastin and factor V are not present pre-formed in the plasma but are themselves generated during the complex process of coagulation from prothromboplastin in conjunction with calcium ions, a labile factor V (plasma thromboplastin accelerator) and a stable factor VII (convertin). Prothromboplastin (also called tissue thromboplastin) is found in saline extracts of many tissues, especially brain, placenta, testis and lung and it is liberated from these and other tissues as the result of injury. Prothromboplastin is a lipo-protein complex which does not exist pre-formed in the plasma. It is produced there, thus initiating coagulation, by the disintegration of platelets on contact with a water-wettable surface to yield a lipoid substance which combines, in the presence of calcium, with two protein factors both normally found in plasma (antihæmophilic globulin and Christmas factor) to give prothromboplastin.

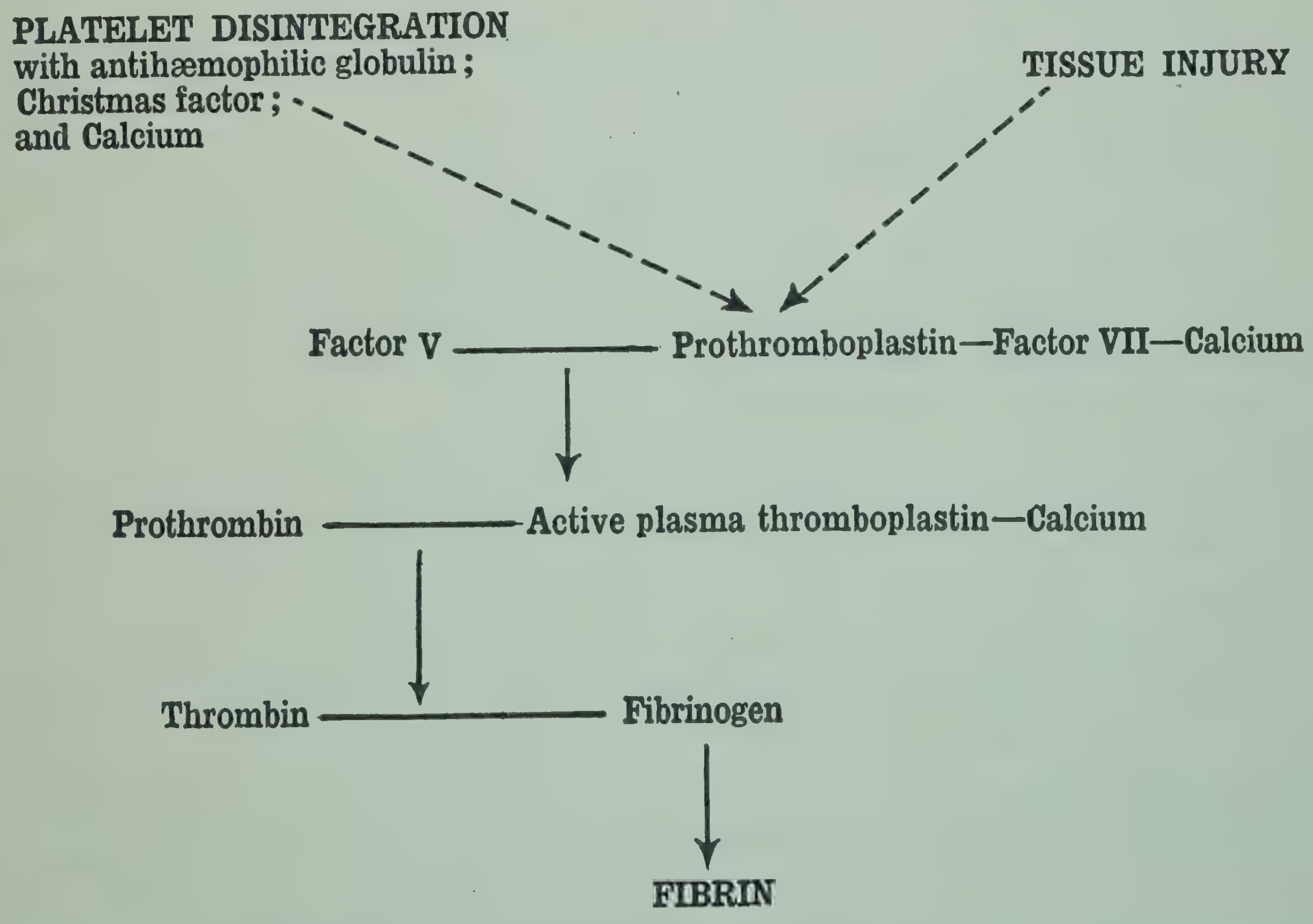

PLATELET DISINTEGRATION
with antihæmophilic globulin;
Christmas factor;
and Calcium
TISSUE INJURY
Factor V
Prothromboplastin—Factor VII—Calcium
Prothrombin
Active plasma thromboplastin—Calcium
Thrombin
Fibrinogen
FIBRIN

Fig. 23, 5

The stages of the process of coagulation in summary (simplified from R. Biggs & R. G. Macfarlane (1953). *Blood Coagulation.* Oxford: Blackwell).

It has been described on page 361 how the efficiency of the whole complex series of reactions can be assessed in terms of the time taken for complete clotting to occur (*coagulation time*). The efficiency of the later stages of the process, namely the conversion of prothrombin to thrombin and the reaction of the latter with fibrinogen to form fibrin can be measured by the *one-stage prothrombin time.* Here a potent preparation of thromboplastin, prepared from brain, is added to recalcified oxalated plasma. Normally clotting occurs within 12 seconds. This test is extensively used in the control of treatment with anticoagulants like dicoumarol for the result may be abnormal while the overall coagulation time is still within normal limits. In spite of its name this test does not measure the concentration of prothrombin in the plasma; the result is probably largely dependent on factors V and VII. It is to be understood that the one-stage prothrombin time may be normal even though there are gross deficiencies of the factors involved in the production of active plasma thromboplastin, for example in hæmophilia.

When normal human blood clots almost all the contained prothrombin has disappeared from the serum within an hour. When, however, the production of active plasma prothrombin or of prothromboplastin is defective there is delay in the prothrombin used up so that at the end of one hour more remains in the serum than normally. By assessing, therefore, the amount of prothrombin remaining in serum one hour after clotting (*prothrombin consumption test*) the adequacy of the reactions leading to the formation of thromboplastin is tested. A more specific and elaborate test for the same purpose is the *thromboplastin generation test* which tests individually the various factors involved and can therefore identify a specific fault, such, for example, as a deficiency of antihæmophilic globulin.

Clot retraction.—It is important to recognize that the mere clotting of blood is insufficient for complete hæmostasis. After the rapid formation of fibrin is complete further changes gradually take place, so that the normal blood clot contracts to about 40 per cent. of its initial volume within twenty-four hours. The threads of fibrin draw more closely together, the texture of the clot becomes firmer and it adheres closely to the containing vessel. For this phenomenon blood platelets are necessary and in their absence hæmostasis may be incomplete. The serum expressed during clot retraction contains much thrombin.

The influence of platelets in coagulation has been studied by Tocantins, who watched under the microscope the coagulation of normal blood and that of slowly clotting hæmophilic blood in a platelet-rich and platelet-free condition. Soon after blood is shed fibrin needles appear, to which the platelets cling like raindrops on a telegraph wire. At the intersections of the needles the platelets form large knots, presumably strengthening the fibrin framework, which then becomes bent, twisted and shortened. This process is probably responsible for clot retraction (Fig. 23, 6).

Hæmophilia is characterized by the development of large hæmorrhages into the tissues and by bleeding into the joints; the coagulation time is greatly prolonged—up to several hours instead of several minutes. Between attacks of bleeding, however, the coagulation time may be nearly normal and the patient relatively well. A small quantity of normal blood or of plasma added to hæmophilic blood brings about rapid clotting. The one-stage prothrombin time is always normal, suggesting that the fundamental defect is earlier in the coagulation process. The thromboplastin generation test shows that antihæmophilic globulin is lacking in the plasma.

During a bleeding phase hæmorrhage is often speedily arrested by blood transfusion which may reduce the coagulation time to normal in a few minutes. The blood must be fresh, since antihæmophilic globulin is labile, but the effect is often very temporary, lasting some 24 hours only. Local bleeding, as from the nose or gums, can often be arrested by the application of preparations containing thromboplastin. The inheritance of hæmophilia is discussed in Chapter 54. A similar disorder, Christmas disease, is the result, not of lack of antihæmophilic globulin, but of another factor (Christmas factor, plasma thromboplastin component).

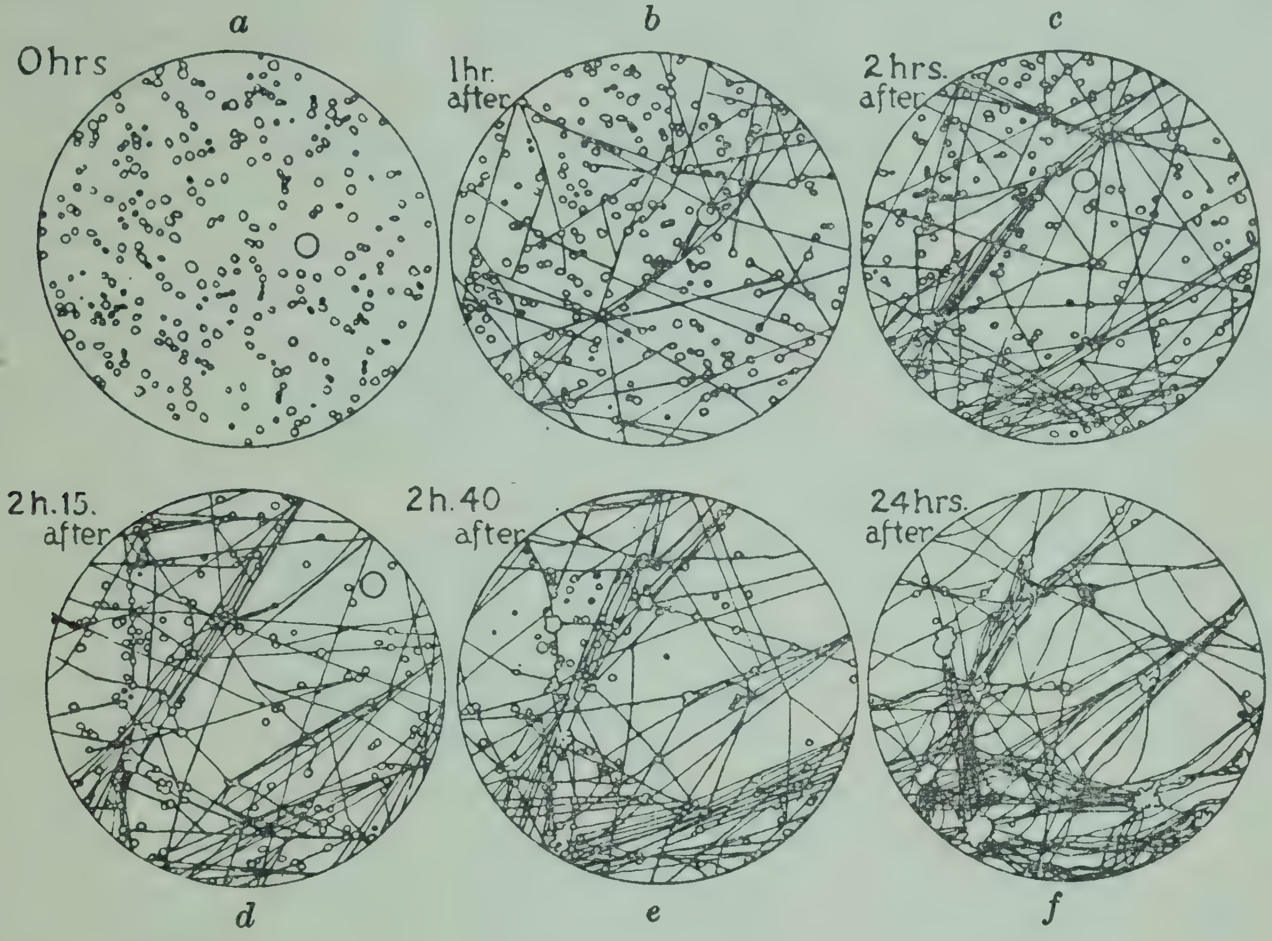


FIG. 23, 6

Drawings of a microscope field of platelet-rich plasma from a person suffering from hæmophilia. The slow rate of clotting in this condition allows detailed observation of the process.

(*a*) Initial appearance showing platelets only.
(*b*) One hour later; fibrin needles have appeared and the platelets are moving towards the needles.
(*c*) Two hours later; nearly all the platelets are attached to the fibrin needles; knots are forming at the intersections of needles.
(*d*), (*e*) The platelets have now formed into groups and the fibrin needles are bent and twisted.
(*f*) The needles have formed a twisted, knotted, intertwined mass which has contracted.

In normal persons the fibrin needles are formed much more quickly and the needles are shorter, thinner and more closely packed (L. M. Tocantins (1935), *Amer. J. Physiol.* **114**, 711).

Fibrinolysis.—If clotted blood is kept sterile the clot remains intact for several weeks. Under certain conditions, however, the clot breaks up and disappears, a phenomenon known as *fibrinolysis*. Normal plasma contains an enzyme, *plasmin*, responsible for the lysis of fibrin and fibrinogen, and also an antiproteolytic enzyme, *antiplasmin*, associated with the globulin fraction of the plasma proteins. There is sufficient plasmin in the blood to destroy all its fibrinogen in a short time but the antiplasmin normally present is more than sufficient to inhibit all the plasmin and fibrinolysis is prevented. Blood removed after severe exercise exhibits fibrinolysis but blood obtained after half an hour's rest shows no fibrinolysis. The phenomenon can be demonstrated in blood removed ten to twenty minutes after the subcutaneous injection of 1 ml. of 1: 1000 adrenaline and after surgical operations, accidents, fear and exercise. All these conditions are associated with the release of adrenaline into the blood but the precise relationship between adrenaline and fibrinolysis is not known.

Hæmostasis.—While disturbances in blood coagulation can themselves be the cause of bleeding diseases, the clotting of blood is by no means the only, and probably not even the most important, process involved in hæmostasis. Patients with the rare condition afibrinogenæmia, in which the blood is completely incoagulable because of the absence of fibrinogen, are able to lead a relatively normal life, being less severely incapacitated than many hæmophiliacs in whom the coagulation time may be no longer than forty-five minutes. Normal blood when removed from the body coagulates in five to ten minutes. However, measurements of the *bleeding time* made by observing how long bleeding persists from a small stab wound some 4 to 5 mm deep made in the lobe of the ear, or the pulp of a finger, are shorter than the coagulation time. Probably the main factor involved in the immediate arrest of bleeding is constriction of the vessel walls and the formation of plugs of agglutinated platelets; the disintegration of platelets actually liberates a vasoconstrictor material. A few minutes after this clotting of the blood occurs. About half an hour later, when the capillary vessels reopen in response to the accumulation of tissue metabolites, the area has been sealed by a blood clot and hæmostasis is maintained (top part of Fig. 23, 7). Later still the clot retracts to form a firm plug, which finally becomes replaced by fibrous tissue. The fibrin network of the clot may act as a scaffold for the new collagen and the new capillaries. Unless both the vascular contraction and coagulation occur in a normal fashion the hæmostatic mechanism is deranged, as shown in Fig. 23, 7.

When the concentration of circulating blood platelets is unusually low (thrombocytopenia) there is apt to be continuous oozing of blood from small wounds. Here the defect is not, as in hæmophilia for example, the result of defective coagulation, but probably rather of the failure of the minute blood vessels to contract, because of the lack of a vasoconstrictor substance (possibly 5-hydroxytryptamine) liberated from the blood platelets.

If bleeding occurs because of a defect in clotting the best results are achieved by supplying the missing factors. Local application of thrombin or the venom of Russell's viper (containing an active thromboplastin) causes local clotting and hæmostasis if the clots are not washed away by continued flow of blood. Fibrin in the form of sheets or foam has also been used successfully; it acts as a dressing which is absorbed during healing. Any hæmostatic dressing must, of course, be applied with gentle pressure and be left undisturbed.

Thrombosis.—In spite of the factors inhibiting the coagulation of blood within the vessels intravascular clotting does sometimes occur, much more often in veins than in arteries, for in the former, of course, the rate of flow is slower. Venous thrombosis is much commoner in the veins of the legs than elsewhere, particularly in the deep veins of the calves. Prolonged rest in bed and varicosity of the veins, which are both accompanied by further slowing of the blood flow, favour intravascular clotting. In addition to sluggish flow of blood there are certain changes in the blood maximal some eight to twelve days after surgical operations which also favour thrombosis. These are an elevated platelet count, increased stickiness of the platelets and a raised concentration of fibrinogen. The tendency to venous thrombosis is reduced by the administration of the anticoagulants heparin and dicoumarol.

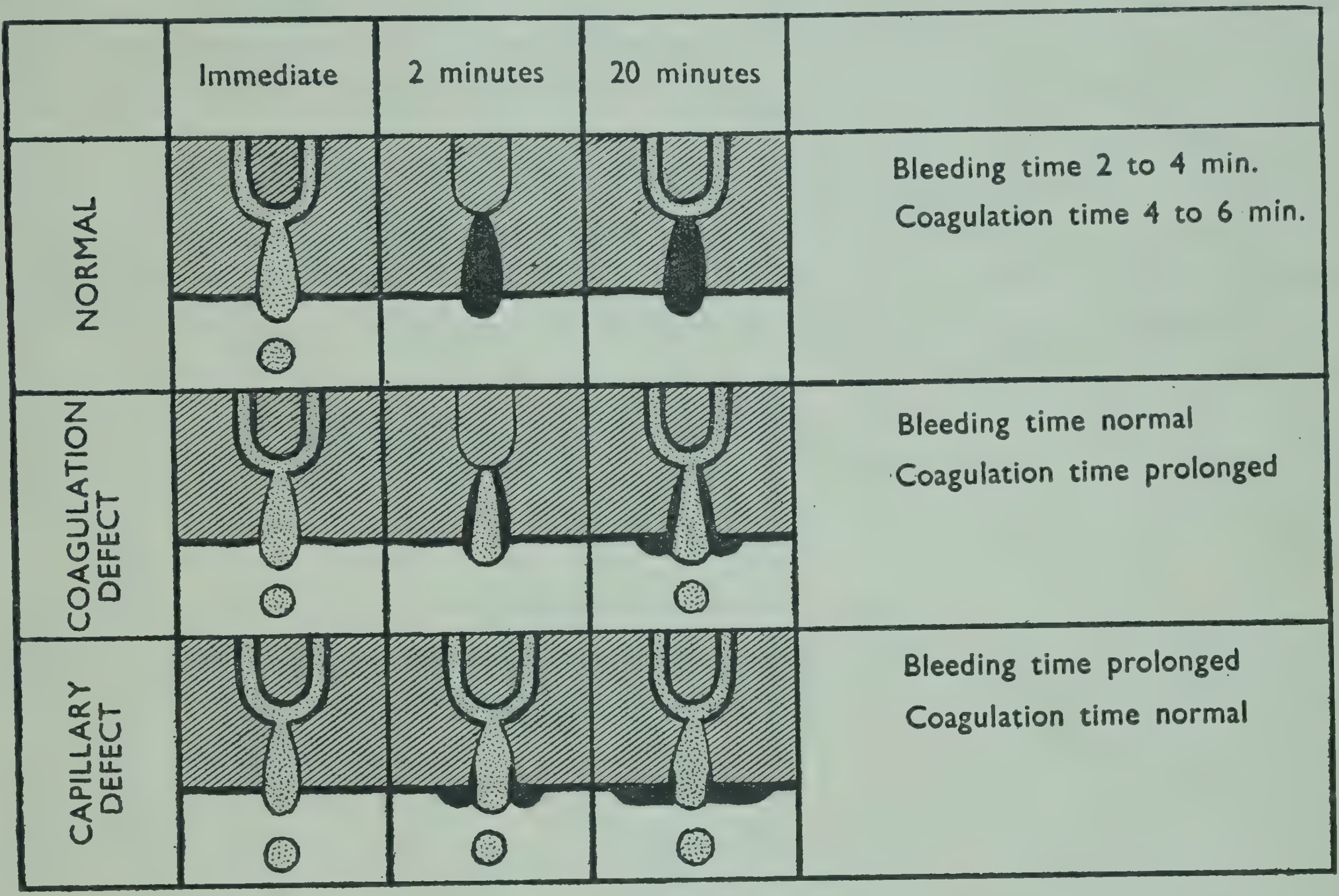


FIG. 23, 7

Diagram illustrating the time relationship of normal capillary contraction, dilatation and blood coagulation, and the two main defects that may occur. A wound of the skin surface is shown in section, injuring a capillary loop. Fluid blood is represented by the dotted areas, blood clot by solid black, and the detached 'drops' indicate active hæmorrhage (R. G. Macfarlane (1944), *Proc. Roy. Soc. Med.* **38**, 400).

THE PLASMA PROTEINS

Blood plasma contains about 9 per cent. of solids, some 7 per cent. being protein in nature. According to the classical view, the plasma proteins are divided into three main types: *fibrinogen, albumin* and *globulin.* Each of these types is actually a group of proteins with similar electrical properties (isoelectric point and surface charge for example). Thus 'plasma albumin' consists of at least two proteins while 'plasma globulin' has been further subdivided into α, β, and γ fractions. It is usually stated that the globulins are precipitated at half saturation of plasma with ammonium sulphate and albumins at full saturation, but the method of fractional precipitation is in fact much less crude than this. Fibrinogen precipitates at 0·25 saturation, γ-globulin at 0·34 saturation, α- and β-globulin at 0·5 saturation and albumin at 0·68 saturation. A similar scheme of fractionation has been devised using sodium sulphate. These fractions, however, are far from precise and much overlapping occurs between fractions precipitated at a particular salt concentration.

The late E. J. Cohn, at Harvard, devised a system for separating the plasma proteins by precipitation with various concentrations of alcohol at temperatures below 0°. At such low temperatures denaturation does not occur and it is possible to obtain plasma proteins in bulk in a relatively pure state. At least 26 pure proteins have been obtained from human plasma by this method. A modification of this procedure is in use at the Lister Institute in London to obtain fibrinogen. Plasma at 0° is treated with ether and fibrinogen is precipitated in a fairly pure state. The fibrinogen is removed and the residual fluid, which is essentially serum, is acidified to pH 5·3 or 5·4 to precipitate prothrombin, which is then collected and dried from the frozen state. The residual fluid, which contains albumin and globulin, may be administered intravenously just like blood plasma.

The most efficient separation of the plasma proteins from one another, however, is achieved by the process of electrophoresis already described in Chapter 5. The Tiselius apparatus may be used for large-scale work but electrophoresis on paper (Fig. 5, 7) is now commonly employed for clinical work on plasma proteins.

The concentration of the various types of protein in normal blood plasma is shown in Table 23, 8. The generally accepted range is shown in the second column, while the third and fourth columns show average figures obtained from large numbers of individuals by Cohn and Luetscher respectively. These authorities used different methods and their results differ quite considerably.

TABLE 23, 8

Plasma Protein Concentrations

Average values obtained by Cohn and by Luetscher as the result of analyses of pooled human blood plasma by the method of electrophoresis.

	Generally accepted range g. per 100 ml. plasma	Cohn g. per 100 ml. plasma	Luetscher g. per 100 ml. plasma
Total protein . .	5·5 to 7·6	6·58	6·50
Fibrinogen . . .	0·2 to 0·4	0·31	0·37
Albumin . . .	3·3 to 5·6	3·26	4·06
α-globulin . . .	..	1·01	0·46
β- ,, . . .	..	1·26	0·86
γ- ,, . . .	..	0·74	0·75
Total globulins excluding fibrinogen .	1·6 to 3·1	3·01	2·07
Albumin/globulin ratio	1·5 : 1 to 2·5 : 1	1·08 : 1	1·94 : 1

E. J. Cohn *et al.* (1946), *J. Amer. Chem. Soc.*, **68**, 459. A. J. Luetscher (1940), *J. clin. Invest.* **19**, 313.

Fibrinogen (molecular weight 330,000) is essentially a globulin which is unusually easy to precipitate. It is readily coagulated by heat and can be precipitated by 0·25 saturation with ammonium sulphate. Fibrinogen is the precursor of the fibrin of the blood clot and there is good evidence that fibrinogen is formed almost exclusively in the liver. Solutions of fibrinogen are some six times as viscous as albumin solutions of the same concentration and the viscosity of plasma is largely due to fibrinogen. It can be prepared in bulk in a relatively pure state either by the alcohol precipitation method of Cohn or by the ether precipitation method used at the Lister Institute. By allowing purified fibrinogen to react with thrombin it is possible to prepare fibrin films in a variety of thicknesses of different mechanical strengths. These films can be sterilized by heat. Heating also modifies their mechanical and chemical properties and renders them suitable for use as substitutes for dura mater in operations on the brain. Fibrin may also be prepared in the form of seamless tubing for the repair of blood vessels. When fibrinogen solution is beaten vigorously in a mixer and treated with thrombin a porous foamy rigid clot is formed, which can be cut into pieces, dried from the frozen state, and finally sterilized by heat. This fibrin foam is used in neurosurgery as a hæmostatic. Concentrated solutions of fibrinogen are used surgically for joining the ends of severed nerves; the solution is poured into the gap between the ends and is then made to clot by the addition of thrombin so that a protein joint is formed of material which is not foreign to the body.

The *Globulins*, designated α-, β- and γ-globulins, cover a wide range of molecular weights from 90,000 to 1,300,000. They are produced mainly in the liver but γ-globulins are formed in reticulo-endothelial tissue and in macrophages and lymphocytes.

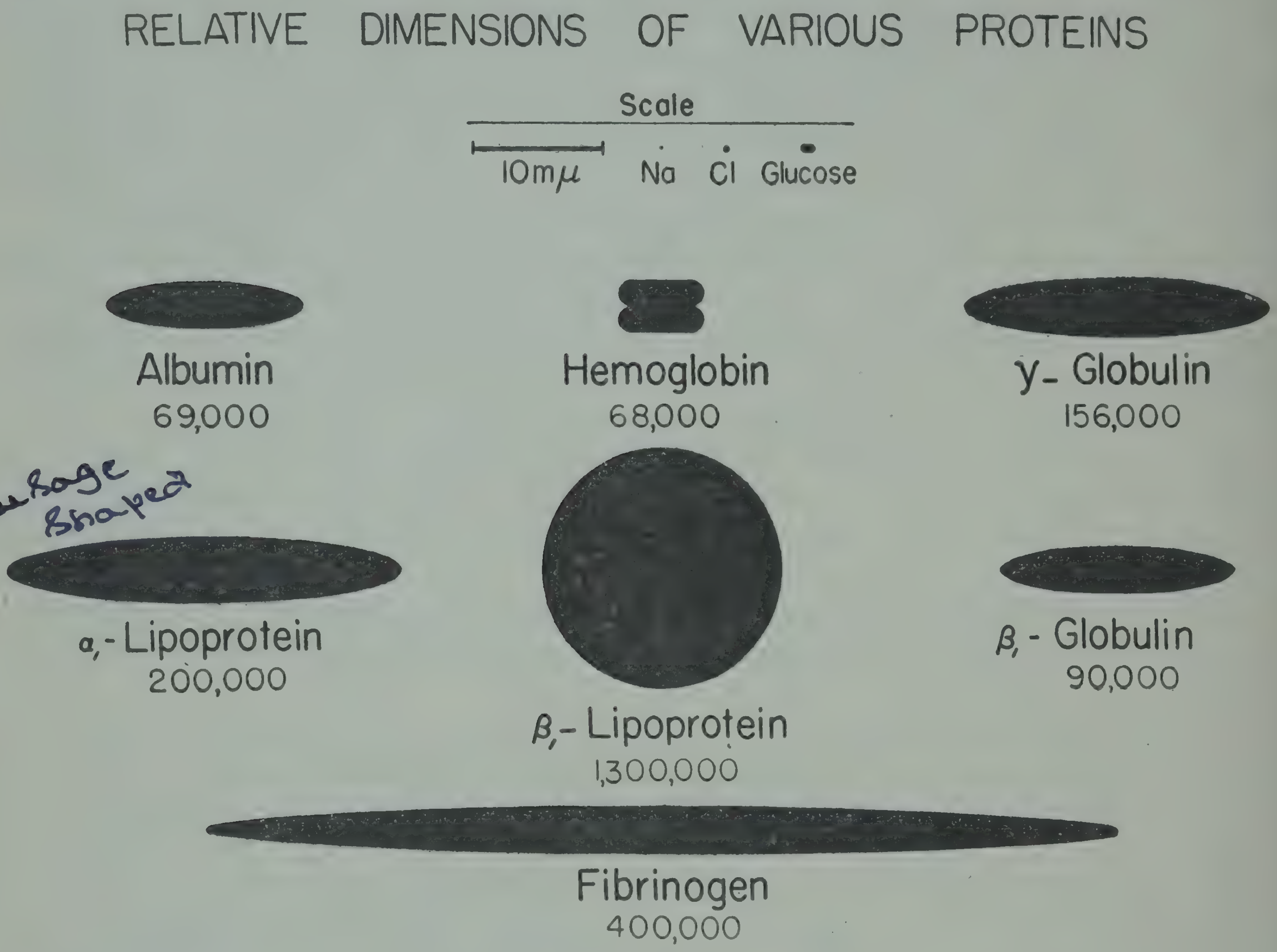
α-globulins include a fraction which combines with bilirubin and another fraction, the α-lipoprotein of the plasma, responsible in part for the carriage of lipids and steroids.

Among the β-globulins are the β-lipoprotein, a very high molecular weight protein responsible for lipid, steroid and carotene carriage, and also a globulin which binds iron and copper for transport in the plasma. Prothrombin is a β-globulin. All known antibodies are included in the γ-globulin fraction. Cases have been described in which there was virtual absence of γ-globulins from the blood without alteration of the other plasma proteins (agammaglobulinæmia). Such patients are unable to produce antibodies and are therefore very liable to recurrent infections which readily prove fatal.

There is a wide field for the use of solutions of gamma globulin for conferring passive immunity. For example, the gamma globulin fraction of the plasma protein of a person who has recently had measles contains a high concentration of antibody to the virus of the disease. Given in the form of 'convalescent serum' or of a preparation of gamma globulins it may abort, or prevent, an attack of measles in a recipient. Such preparations are now beginning to be used in a number of virus diseases in man, in poliomyelitis for example, and in dogs for distemper.

The *Albumins* (molecular weight about 68,000) are the most abundant of the plasma proteins and have the highest electrophoretic mobility. The albumin fraction is formed mainly, but not exclusively, in the liver. Concentrated solutions of human albumin are used instead of blood plasma for transfusion to increase the plasma protein concentration, and therefore the osmotic pressure, with the idea of attracting œdema fluid into the vascular compartment of the extracellular space and so facilitating its excretion. Such measures are not usually very successful because in the circumstances the transfused albumin rapidly leaves the circulation.

Osmotic Pressure of Plasma Protein.—The plasma proteins play an indispensable role in the osmotic equilibrium which is maintained between the blood and the tissue fluids. Albumin is the most important factor in this respect, partly because it is the most abundant protein and partly because it can, owing to its lower molecular weight, exert an osmotic pressure two or three times that of globulin per unit weight. Actually plasma albumin at normal concentrations is responsible for about 80 per cent. of the total osmotic pressure of the plasma proteins. Although only about 25 mm of mercury the plasma protein osmotic pressure is very important in preserving the fluid balance between blood and tissues.


RELATIVE DIMENSIONS OF VARIOUS PROTEINS
Scale
10mμ
Na
Cl
Glucose
Albumin
69,000
Hemoglobin
68,000
γ- Globulin
156,000
α₁- Lipoprotein
200,000
β₁- Globulin
90,000
β₁- Lipoprotein
1,300,000
Fibrinogen
400,000

Sausage
Shaped

FIG. 23, 9

Diagram showing relative dimensions and molecular weights of blood proteins. The molecules of albumin, globulin and fibrinogen are elongated bodies of about the same diameter but of very different lengths according to their molecular weights (E. J. Cohn (1945), *Sci. in Prog.* **4,** 319).

It has recently been found that normal capillary filtrates and normal lymph from the limbs contain small amounts of albumin (molecular weight 68,000) and even of globulin (molecular weight 90,000 to 175,000) and fibrinogen (molecular weight 330,000) but it was difficult to understand how these proteins, especially those of very high molecular weight, could pass through the capillary wall. Cohn has shown that the molecules of plasma proteins are sausage-shaped bodies, almost all of which have nearly the same diameter, namely 3·3 to 3·8 mμ. They vary in length, however, according to their molecular weight (Fig. 23, 9). It appears that, although the majority of pores in the capillary walls measure from 0·7 to

2·0 mμ, there are some at least 3·8 mμ in diameter. All these pores allow simple salts to pass quite freely while the largest of them allow the passage of albumin and even of globulin and fibrinogen provided that these large molecules strike the capillary wall in the correct orientation, that is 'end-on' to the pores. These new facts are sufficient to account for the small amounts of protein which are normally found in capillary filtrates and in urine.

Hypoproteinœmia may be said to exist when the total plasma protein concentration is less than 5·5 g. per 100 ml. It may be caused by insufficient intake of protein, by excessive loss of protein into the urine or extracellular space, or by defective synthesis of protein by a damaged liver. In any given patient the explanation is probably multiple but the result is that the colloid osmotic pressure of the plasma is diminished and the tendency for water to be attracted from the tissues into the blood stream is reduced. Water may therefore accumulate in the tissues; if this is excessive it may become obvious as *œdema.* It was formerly assumed that *famine œdema* was caused by hypoproteinæmia due to insufficient protein in the diet since the plasma protein concentration in this condition is usually below 4 per cent., but results obtained during the last war have shown that famine œdema is not necessarily accompanied by hypoproteinæmia, nor does hypoproteinæmia necessarily result in œdema.

THE FORMATION OF PLASMA PROTEINS

Experiments on dogs by Whipple have shown that the plasma proteins are formed from amino-acids derived from the food. By a technique called *plasmapheresis* he has investigated some of the factors involved in the generation of plasma protein. Plasmapheresis consists in withdrawing a quantity of blood from which the red blood corpuscles, after separation by centrifugation, are washed with isotonic saline and then returned to the dog by intravenous injection. In this way the animal is deprived of plasma proteins without loss of hæmoglobin. If the animal is kept on a standard diet the plasma proteins can be gradually reduced by daily plasmapheresis from the normal level of 6 per cent. to 4 per cent. and can be kept at this lower level.

TABLE 23, 10

Amount of Protein fed	Amount of Plasma Protein formed
100 g. serum protein (ox)	38 g.
100 g. meat (skeletal muscle)	18 g.
100 g. casein	12 g.
100 g. gelatin	5 g.
100 g. zein	0 g.

If now a protein is fed as a supplement to the standard diet without altering the amount of plasmapheresis, the percentage of protein in the plasma rises and the total amount of extra plasma protein manufactured can be estimated if the blood volume is known. Some experimental results are given in Table 23, 10, which is extracted from Whipple's data.

TABLE 23, 11

*Functions of Blood Proteins**

Site in Blood	Principal Function	Principal Protein related to Function	Clinical Use
Red cells which contain 34 per cent. Hb and occupy 45 per cent. of blood volume	Respiratory	Hæmoglobin	Whole blood or red cell transfusion
Plasma which contains 7 per cent. proteins and makes up 55 per cent. of blood volume	Osmotic regulation of blood volume	Albumin	Shock, burns, œdema, hypoproteinæmia
	Blood coagulation	Fibrinogen	Fibrin films as dural substitutes
		Prothrombin, thrombin	Fibrin foam and thrombin as hæmostatic agents
	Immunological	Blood grouping globulins	Blood grouping
		'Complement'	
		γ-globulins	Measles prophylaxis
	Carbohydrate and lipid solution and transport and iron transport	β-globulins	
		α-globulins	
	Regulatory	Fibrinolytic enzyme	
		Phosphatase and other enzymes	
		Hypertensinogen	
		Thyrotropic hormone	
		Gonadotrophic hormone	
		Other protein hormones	

* Modified from E. J. Cohn.

Such experiments indicate that the plasma protein is formed from amino-acids derived from the food and other experiments along similar lines show further that cystine is a key amino-acid in plasma protein regeneration.

Although the proteins of the plasma and of the tissues are continually being replaced by new proteins synthesized from the amino-acids of the food there is good evidence that small amounts of protein can be absorbed directly from the alimentary tract and that they can penetrate cell surfaces and enter cells without prior disintegration into amino-acids. Both animals and human patients can be kept in nitrogen equilibrium (p. 188) by plasma given intravenously. Since the injected protein disappears from the blood very rapidly without causing any rise in blood amino-acids it presumably enters the tissue cells. Such experiments, and others in which protein tagged with isotopic tracers has been used, suggest that there is a continual rapid exchange of protein molecules between the tissues and the blood plasma (p. 339). There is no store of plasma protein as such, but there is a considerable reserve of hæmoglobin- and plasma-protein-building material, so that even after a severe hæmorrhage about half the amount lost can be re-formed in a day. The site of this reserve store is clearly shown by the fact that after severe depletion by plasmapheresis, followed by a period of starvation, plasma protein is regenerated at the expense of tissue protein.

If plasmapheresis ceases the material which would have been converted into the plasma proteins, had plasmapheresis continued, is stored in the body depots. It is much more than can be accounted for by the rise in blood proteins. This material in the depots can be called on again by reinstituting plasmapheresis. Regeneration of plasma protein is retarded during infection. Since there is an ebb and flow between the plasma proteins and the cells of the body, the plasma proteins can be thought of as the mobile exchange material of the total body protein pool. The plasma proteins contribute freely to the formation of hæmoglobin but once globin is incorporated in the red cell as hæmoglobin it is not available for the protein pool.

A summary of the functions and clinical uses of the blood proteins is given in Table 23, 11.

Specific Gravity of Blood.—When a drop of blood is placed in a solution of copper sulphate it is at once covered with a pellicle of copper proteinate and cannot mix with the fluid. By using a series of bottles containing an appropriate range of concentrations of copper sulphate the specific gravity of blood can be estimated very accurately, since the specific gravity of blood is the same as the specific gravity of a fluid in which a drop of blood neither sinks nor rises. If the specific gravity of the serum is measured similarly, the concentration of the plasma proteins can be approximately inferred. This simple method was introduced in 1942 by Van Slyke to help in treating battle casualties. It is justifiable in many cases to assume that the plasma proteins are within normal limits. If this is so the specific gravity of blood depends mainly on its hæmoglobin content, and by the copper sulphate method the hæmoglobin can be estimated very rapidly and with reasonable accuracy.

The normal values for specific gravity are given in the Table 23, 12.

TABLE 23, 12

	Specific Gravity
Whole blood .	1·055 to 1·065
Serum . .	1·028 to 1·032
Corpuscles .	about 1·096

BLOOD VOLUME

The blood volume of an animal can be assessed by collecting all the blood which flows from a cut artery. Bleeding-out by this method yields only part of the total blood in the body but the blood remaining in the tissues can be washed out by perfusing Locke's solution through the vessels. The blood which emerges after the commencement of perfusion is, of course, diluted by the Locke's solution but the true blood volume may be obtained by comparing the depth of colour of the animal's undiluted blood with that of the diluted blood and making an allowance for the extent of the dilution. The blood volume of a dog measured in this way is about 1/13th, or 7·7 per cent., of the body weight.

For clinical purposes the blood volume can be estimated by using a non-toxic dye which escapes very slowly from the blood vessels. A sample of blood from a vein is oxalated and then centrifuged to obtain a sample of plasma free from dye and to find the relative volumes of plasma and corpuscles. A known quantity of a 1 per cent. solution of dye, say 10 ml., is injected intravenously. After about six minutes, during which mixing becomes complete, a second blood sample is obtained and the dye-stained plasma is compared in a colorimeter with a standard dilution of the original dye solution. If the dye-stained plasma matches a 1 in 300 dilution of the original dye solution then the plasma volume must be 300×10 ml., or three litres. If centrifugation of the first blood sample shows that the relative volumes of plasma and corpuscles are 55 and 45, then 3 litres $= \frac{55}{100}$ of the blood volume. The blood volume must, therefore, be $3 \times \frac{100}{55}$, or 5·45 litres.

If further blood samples are obtained it is found that the concentration of the dye in the plasma falls off slowly; by plotting the concentration against the time a curve is obtained which can be extrapolated to zero time (the time of injection of the dye) for calculation of the true plasma volume.

The following results have been obtained by using the blue azo-dye T-1824 (Evans Blue):

	Males	*Females*
Average plasma volume	2·95 litres	2·28 litres
Average total blood volume	5·34 litres	3·80 litres
Blood volume per kg. body weight	78 ml.	66 ml.
Blood volume per sq. metre	3·1 litres	...

The blood volume is higher in muscular active individuals and is related to the height and weight (that is, the surface area) of the individual thus,

$$\text{Blood volume in ml.} = 43W + 52H - 6250$$
$$\text{Plasma volume in ml.} = 20W + 29H - 3250$$

where W = weight in kg. and H = height in cm. At birth the blood volume is about 300 ml., which is rather more than would be expected from the data given above; in the first months of life the blood volume is about

50 per cent. greater, and at one year about 10 per cent. greater, than would be calculated from the body weight. Although the volume of plasma per square metre is similar in men and women the total blood volume of men is greater by about 7·5 per cent. because of a larger red cell mass.

The dye method makes the assumption that the ratio of plasma to corpuscles is the same all over the body but the following recent experiments show that this is not true. By giving a dog a transfusion of red cells containing radioactive iron ^{55}Fe and of albumin iodinated with radioactive iodine ^{131}I it was possible, by calculations similar to those described above, to measure independently the total red cell volume and the total plasma volume. The animal was then killed and the radioactivity of the minced organs gave a measure of the relative proportion of plasma and red cells in the small vessels (Table 23, 13). The low figures for certain small vessels may be due to 'plasma skimming'; many minute vessels may contain only plasma. The high proportion of red cells in the blood in the spleen may be associated in some way with the sluggish flow in the splenic pulp. Because of this variation in the proportion of plasma and red cells in different parts of the circulation the ratio of total red cell volume measured by radioactive iron to the total red cell volume obtained by the dye method is about 0·85. This means that the dye method gives results for the volume of red cells that are 15 per cent. too high.

TABLE 23, 13

Tissue	Hæmatocrit Reading. Volume of red cells as a percentage of volume of whole blood (dog)
Arterial blood	43
Small vessels of—	
Spleen	80
Liver	40
Lungs	35
Heart and skeletal muscle . .	20 to 25
Kidney, gastro-intestinal tract and brain	15 to 20

These results were obtained in the dog by means of radioactive iron and iodine as described in the text.

From J. G. Gibson, A. M. Seligman, W. C. Peacock, J. C. Aub, J. Fine & R. D. Evans (1946), *J. clin. Invest.* **25**, 848.

Only one-sixth of the entire blood volume is found in the small vessels at any one instant, the remainder being in the large vessels and heart. The blood content of various organs is given in Table 23, 14, and the data of Tables 23, 13, and 23, 14, are combined schematically in Fig. 23, 15.

If a dog is given two successive transfusions of red cells, in the first of which the cells are labelled with ^{55}Fe and in the second with ^{59}Fe, it is found that the two lots of red cells are quickly mixed. All the red cells, therefore, are in active circulation in the dog and there are no reserves of red cells which do not enter the circulation. The rise in the hæmatocrit reading (that is the percentage of red cells) after giving adrenaline was formerly ascribed to the outpouring of red cells from the contracting spleen, but recent studies with radioactive iron show that there is no alteration in the total mass of red cells but merely some redistribution which results from the vasoconstriction.

Antigens and Antibodies.—Before going on to discuss blood groups it is necessary to describe the body's reaction to the introduction of foreign protein. This topic has already been mentioned briefly on p. 55. The phenomenon was first studied in connexion with the immunity which develops, for example, after infection with the bacilli of diphtheria. If the patient survives the disease his serum is found to contain *antitoxins* which can neutralize the *toxins* produced by the bacilli. Nearly all proteins (insulin is a notable exception) act as *antigens* which call forth *antibodies* provided that they are injected into an animal which does not already possess the protein in its tissues, that is, the antigens must be foreign to the animal. The antibodies which are formed are highly specific. By injecting human serum into one rabbit, ox serum into another, and so on, we can obtain rabbit sera which can be used in a precipitation test to identify the animal from which a serum comes even if only a minute quantity of protein, such as might be obtained from a blood stain, is available for test. Anti-human rabbit serum mixed with human serum causes a precipitate to appear but no other antiserum has this effect. If in much the same way red blood cells are injected into an animal of

TABLE 23, 14

The Distribution of Blood in various organs of the Dog

	ml. per g. tissue	Percentage of total blood volume contained in various tissues
Spleen	0·4	1 to 2
Liver	0·2	5
Lungs	0·2	1 to 2
Heart	0·06	< 1
Skeletal muscle	0·02	7
Kidney	0·2	1 to 2
Gastro-intestinal tract . .	0·06	1 to 2
Brain	0·02	< 1
Large vessels and heart . .	. .	83

From J. G. Gibson *et al.* (1946), *J. clin. Invest.* **25**, 848. See also E. J. Delorme, A. I. S. Macpherson, S. R. Mukherjee & S. Rowlands (1951), *Quart. J. exp. Physiol.* **36**, 219.

another species, specific *hæmolysins* are formed. Thus if monkey red cells are injected into a rabbit, the rabbit's serum is found after some days to possess hæmolytic properties which are revealed by mixing monkey corpuscles and the rabbit's serum *in vitro.* The cells are first agglutinated and then lysed so that hæmoglobin is released and the serum becomes a transparent red colour. Corpuscles from any other species mixed with this rabbit serum are unaffected and simply sink to the bottom of the test-tube.

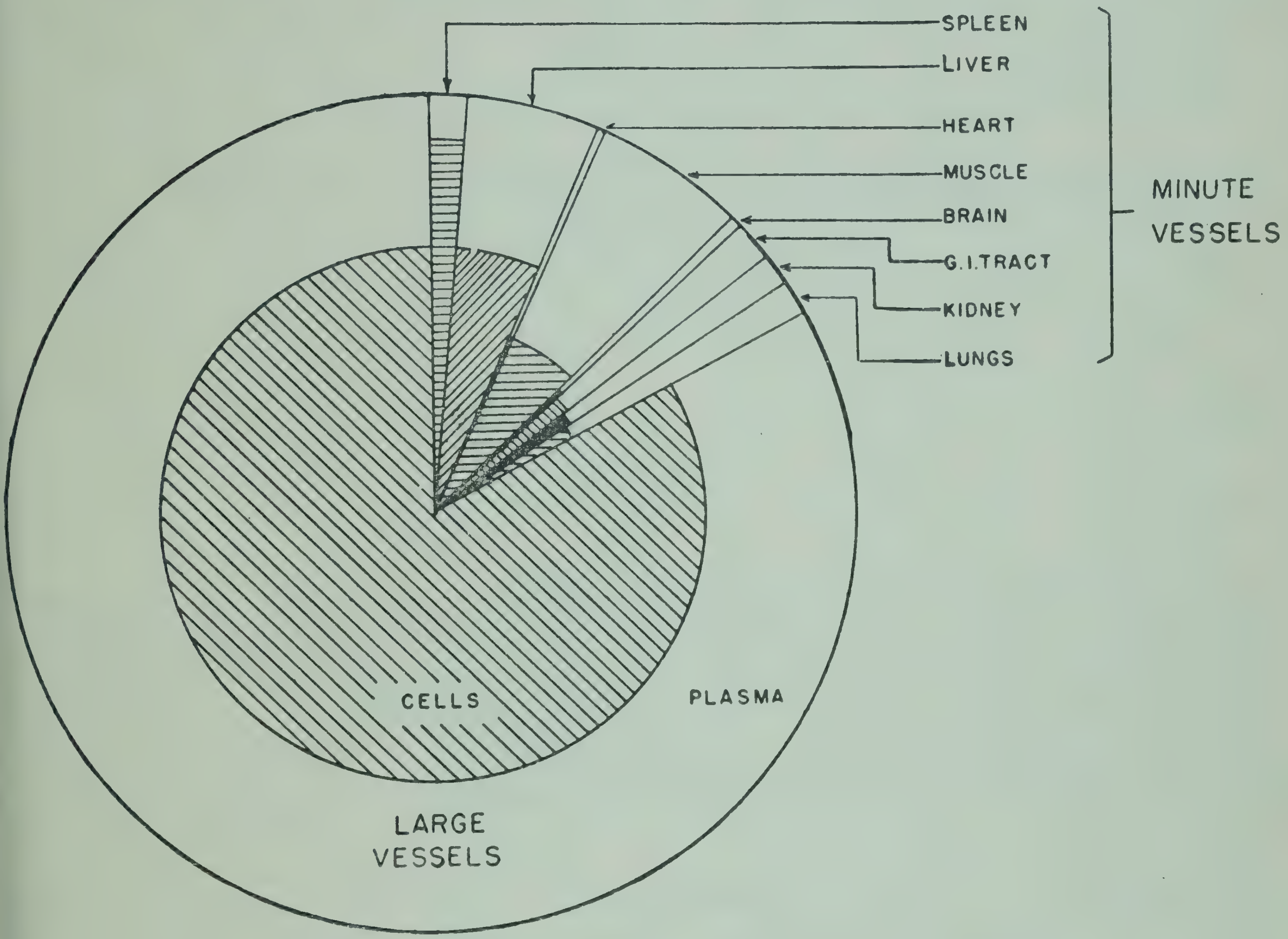


FIG. 23, 15

The relative distribution of whole blood, cells and plasma in the large and minute vessels of the dog (J. G. Gibson *et al.* (1946), *J. clin. Invest.* **25**, 848).

BLOOD GROUPS

Not long after the discovery of the circulation of the blood it occurred to several people that it might be possible to transfuse blood; indeed this is a logical deduction from that discovery. The first experiments on dog to dog transfusion were done about 1665 at Oxford. Pepys describes in 1666 and 1667 experiments in transfusing sheep blood into a man. The first transfusion of human blood was made in 1818, but so many transfusions were followed by death that the procedure fell into disrepute. It was only after further deaths had resulted from attempts to transfuse battle casualties in the early part of the 1914–18 War that the work of Landsteiner in 1900 on blood groups was brought to the notice of surgeons. With the recognition of the compatibilities and incompatibilities of the blood of different persons blood transfusion became relatively safe and when in 1914 sodium citrate, a non-toxic anticoagulant, came into use blood transfusion became quite a simple procedure which saved the lives of many casualties. Occasionally, however, a blood transfusion was followed by fatal results which could not be explained until the Second World War.

When the blood of two individuals is mixed *in vitro* either there is no change and the bloods are said to be compatible, or there is agglutination or clumping of the corpuscles followed by lysis and the two bloods are said to be incompatible. It is now known that there are four main groups of blood determined by the presence of two *agglutinogens* (antigens) A and B in the red cells and two *agglutinins* (antibodies) α, or anti-A, and β, or anti-B, in the serum. The characteristics of the four main blood groups in respect of these are shown in Table 23, 16. If agglutinogen A meets agglutinin α, agglutination occurs: similarly with B and β. The result of mixing all the possible combinations of the four types of cells and four kinds of sera is shown in Table 23, 17, where + indicates agglutination and 0 its absence. Although this table looks complicated it can readily be constructed if four simple principles are recognized: (1) The

TABLE 23, 16

Group	Agglutinogens present in Red Cells	Agglutinins present in Serum or Plasma	Percentage of Persons in Group	
			Scotland	London
AB	AB	O	3	3
A	A	β	35	45
B	B	α	10	10
O	O	$\alpha\beta$	52	42

TABLE 23, 17

Serum of Group	Red Cells of Group			
	AB	A	B	O
AB	0	0	0	0
A	+	0	+	0
B	+	+	0	0
O	+	+	+	0

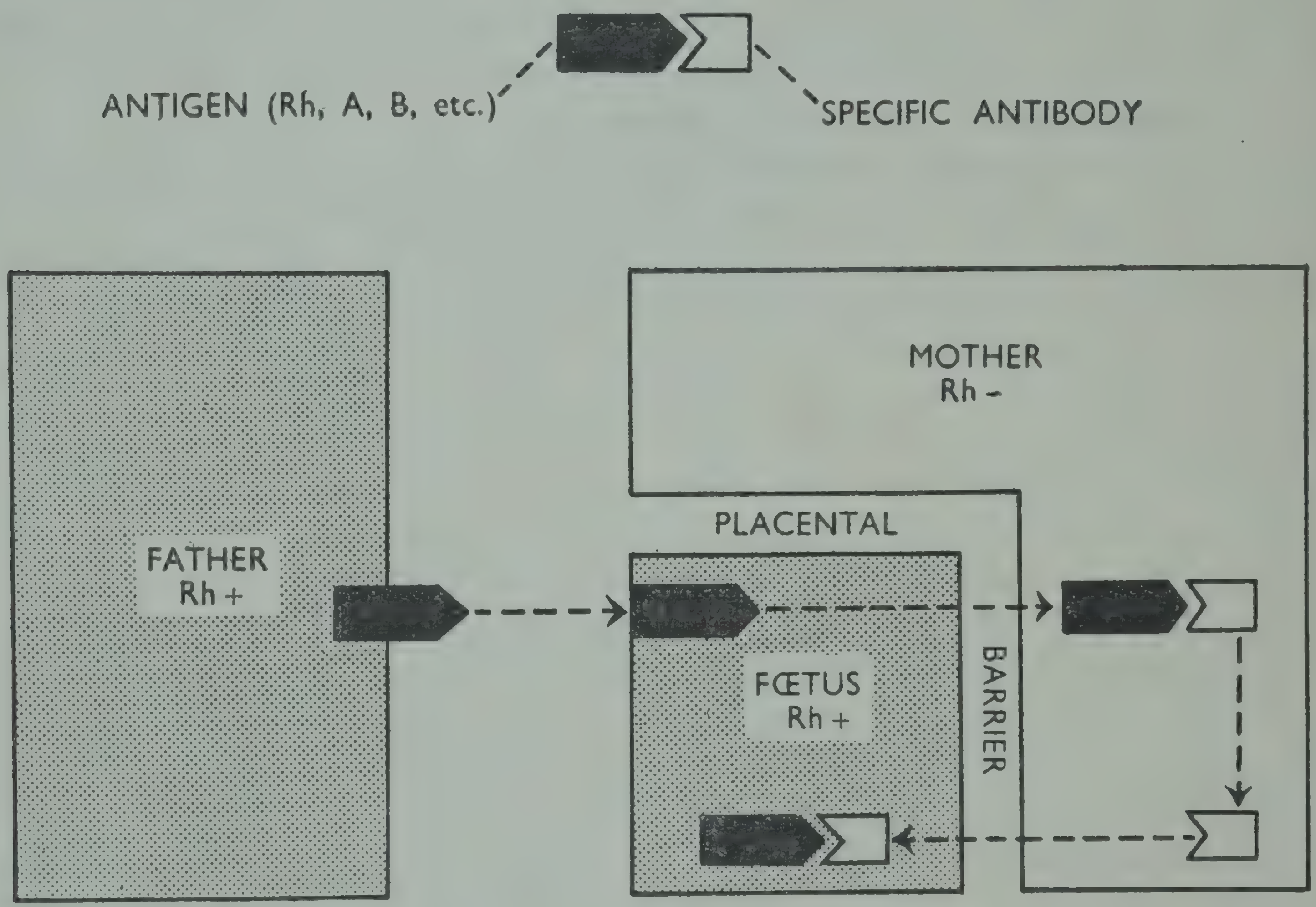
cells of one group are compatible with the serum of the same group. (2) Since the serum of group AB contains no agglutinin the red cells of any group are unaffected by it. (3) Since the red cells of group O contain no agglutinogen they are not agglutinated by any of the sera. (4) On all other occasions there is agglutination.

Inspection of the two middle horizontal rows of Table 23, 17, shows that if serum A and serum B are available any sample of red cells can be allocated to its group, for if the corpuscles are agglutinated by both sera they are group AB; if by B alone they are group A; if by A alone they are group B; if no agglutination occurs the corpuscles are group O. Incompatibility of the mixture of serum and cells is revealed by agglutination, that is by the appearance of granules like brick dust usually visible to the naked eye. Agglutination usually occurs within a minute or two but may be delayed for as long as ten minutes.

Since plasma usually contains only a low titre (low concentration) of agglutinins the effect of the donor's plasma on the recipient's red cells may as a rule be neglected. In any case the risk of hæmolysis of the recipient's cells by the donor's plasma is further reduced by the dilution of the donor's plasma as it mixes with the recipient's plasma. However, when large transfusions, say of more than a litre, of group O blood are given to patients of other groups, then the concentration of anti-A and anti-B agglutinins may be raised to dangerous levels, such that subsequent transfusion of even a small amount of group A or group B blood—that is blood of the same group as the patient—would lead to a severe or even fatal hæmolytic reaction. A much greater hazard, of course, is agglutination of the donor's corpuscles by the recipient's plasma, and it is in this way that the gravest consequences can occur. Such severe hæmolytic reactions, which may cause irreparable damage to the renal tubules, should be avoidable if both donor and recipient are carefully grouped and if a direct *in vitro* test for compatibility be made before the transfusion (see below). However, the safest way of avoiding transfusion reactions is never to transfuse blood of a group other than that of the recipient. Plasma alone is often transfused, especially in the treatment of burns; if pooled plasma (that is, plasma obtained from many persons) is used the agglutinin content can be neglected, since any chance high-titre sample is diluted and its agglutinins are at least partly absorbed by the agglutinogen present in small quantity in plasma.

The earlier statement that the red cells of group O contained no agglutinogen is not quite true, because repeated transfusions of group O cells are capable of stimulating the production of anti-O (or α_2) agglutinin in the plasma of the recipients. Subsequent transfusions with group O blood may then give rise to intravascular agglutination and hæmolysis. An additional complication is the fact that both groups A and B contain subgroups. However, in a first transfusion, subgroup incompatibility is negligible, although repeated transfusions with cells of the same subgroup —a most unlikely event—could conceivably give rise to the development of antibodies and thus lead to a hæmolytic reaction. These and other transfusion hazards can usually be avoided by a direct compatibility test, that is, by mixing a 1 in 20 dilution of the donor's cells with a sample of the recipient's serum. In the absence of agglutination a transfusion reaction due to incompatibility is unlikely to occur, but again it must be emphasized that by far the wisest course is to transfuse only blood of the same group as that of the recipient. The inheritance of blood groups is discussed in Chap. 54.

Another factor quite different from the A, B, O factors is the Rh, or *rhesus factor*. It was originally discovered that, by injecting red cells from a rhesus monkey into a rabbit or guinea-pig, antibodies are developed which agglutinate not only the cells of the rhesus monkey, but also the cells of 85 per cent. of persons. These persons are, therefore, termed Rh-positive, the remaining 15 per cent. being Rh-negative. (It is assumed in the description of the transfusions given below that the bloods show A, B, O compatibility.) If Rh-positive blood is given to an Rh-negative recipient no obvious reaction follows but the Rh-positive corpuscles stimulate the production of anti-Rh agglutinins. Where the titre (concentration) of these is high enough a second transfusion of Rh-positive blood is likely to produce a hæmolytic reaction. In a first blood transfusion the Rh-factor is not of any importance, except in pregnant women, since anti-Rh agglutinins do not occur otherwise. A pregnant Rh-negative woman can produce Rh-antibodies in her serum, as explained below, so that a hæmolytic disaster may occur with the first blood transfusion. The blood groups, including the Rh-factor, are inherited according to the Mendelian laws, so that an Rh-negative mother may, if the father of her child is Rh-positive, have an Rh-positive child. If there is a leakage of



Rh ISOIMMUNIZATION

FIG. 23, 18

Diagram to illustrate the formation of Rh antibodies by an Rh − mother. Not every Rh + father has an Rh + child. If the father is homozygous all his children will be Rh +; if he is heterozygous then there are equal chances of Rh + and Rh − children (S. H. Polayes and T. McNally (1948), *Amer. J. clin. Path.* **18,** 379).

Rh-factor from the child to the mother, across the placental barrier, the mother develops anti-Rh agglutinins (Fig. 23, 18). Thus if there is a high anti-Rh titre a first transfusion of Rh-positive blood to the mother may lead to disaster, although a transfusion with Rh-negative blood is quite safe.

There is another consequence of the carrying of an Rh-positive child by an Rh-negative mother (Fig. 23, 18). The anti-Rh agglutinins formed in the mother's blood may pass back through the placenta into the fœtus and damage the fœtal red cells. The fœtus replies to this blood destruction by active hæmopoiesis. The degree of damage to the fœtus varies with the extent of the hæmolysis—it may not survive or it may survive with anæmia and jaundice. The child in such cases should be transfused after birth with Rh-negative group O blood or Rh-negative blood of its own group. Recent work has shown that the Rh-factor is not a single one but comprises probably eight different antigens.

The effects of Rh incompatibility do not occur in all cases where the child is Rh-positive and the mother Rh-negative, possibly because of variation in placental permeability to the antibodies. Even in families in which hæmolytic disease is manifest it is often confined to, or occurs more severely in, the later members of the family as the result of the increased titre of agglutinins in the maternal blood. If the father is homozygous (Rh +, Rh +) then all his children will be Rh + and it is likely that the Rh − mother will acquire fresh antibody with each pregnancy. When the father is heterozygous (Rh +, Rh −) then there are equal chances of Rh + and Rh − children and the danger of hæmolytic disease in the family is correspondingly less (see Chapter 54).

Hæmagglutinins and Hæmagglutinogens.—The hæmagglutinins occur in the globulin fraction of the serum and in small amount in the body fluids, such as milk or lymph.

The hæmagglutinogens A, B and O are present in all the cells of the body as well as in the red cells. In most people, called *secretors*, these substances are water-soluble and occur in the tissue fluids as well as in the saliva, gastric mucin and ovarian cysts. In *non-secretors* the agglutinogens are confined to the cells. The group A substance contains a polysaccharide-amino-acid complex. Several other groups (not all involved in blood-transfusion reactions) have been discovered, for example M, N, P, S, Kell, and new ones will probably be added to the list. The blood groups are all inherited according to Mendelian laws and remain fixed in any individual throughout life. The number of combinations and permutations of these groups with the A, B, O and Rh groups and their subgroups is enormous, so that if further groups and subgroups are identified it may one day be possible to establish an individual's identity by his blood groups, just as can be done now with fingerprints.

Transfusion.—Blood for transfusion is obtained by putting a needle, with sterile precautions, into a vein of a donor and is collected in a bottle containing a solution of sodium citrate and glucose. The blood can be used at once or it may be stored at 4° C until required. The leucocytes degenerate and lose their phagocytic power within twelve to twenty-four hours. The platelets survive only a few hours and certain labile plasma factors also disappear very early. From the point of view of red cell survival and the carriage of oxygen, stored blood is as good as fresh blood and it may be used up to three weeks. At the end of this period the plasma is usually drawn off because it is of considerable therapeutic value. Fluid plasma keeps at 4° C for at least two years and if dried and stored dry it keeps indefinitely.

When the blood is to be used for therapeutic purposes it is first filtered and then passed slowly through a needle into a vein of the recipient. In children, in whom veins are small, blood can be successfully transfused into the marrow cavity of the tibia.

REFERENCES

Astrup, T. (1950). Blood clotting and related processes. *Advanc. Enzymol.* **10,** 1–49.

Biggs, R. & MacFarlane, R. G. (1953). *Blood Coagulation.* Oxford: Blackwell.

Budtz-Olsen, O. E. (1951). *Clot Retraction.* Oxford: Blackwell.

Edsall, J. T. (1947). The Plasma Proteins and their Fractionation. *Advanc. in Prot. Chem.* **3,** 384–479.

Gutman, A. B. (1948). The plasma proteins in disease. *Advanc. in Prot. Chem.* **4,** 155–250.

Lutz, B. R. (1951). Intravascular agglutination of the formed elements of the blood. *Physiol. Rev.* **31,** 107–130.

Madden, S. C. & Whipple, G. H. (1940). The plasma proteins. *Physiol. Rev.* **20,** 194–217.

Medical Research Council (1943). The determination of blood groups. M.R.C. War Mem. No. 9. Reprinted 1951. London: H.M.S.O.

Mollison, P. L. (1951). *Blood Transfusion in Clinical Medicine.* Oxford: Blackwell.

Mollison, P. L., Mourant, A. E. & Race, R. R. (1954). The Rh groups and their clinical effects. *Spec. Rep. Ser. med. Res. Coun., Lond.*, No. 27. London: H.M.S.O.

Mourant, A. E. (1954). *The Distribution of Human Blood Groups.* Oxford: Blackwell.

Quick, A. J. (1951). *The Physiology and Pathology of Hemostasis.* London: Kimpton.

Race, R. R. & Sanger, R. (1954). *Blood Groups in Man.* Second edn. Oxford: Blackwell.

Tullis, J. L. (1953). *Blood Cells and Plasma Proteins: their State in Nature.* New York: Academic Press.

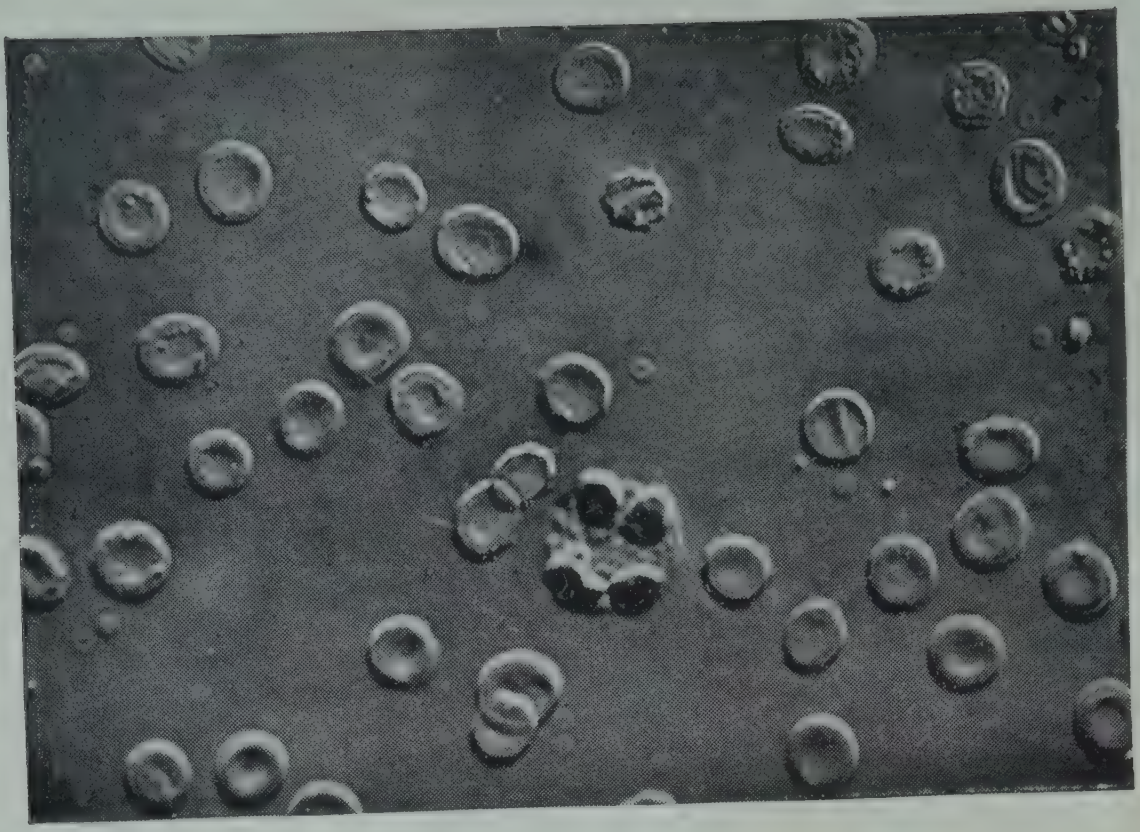


FIG. 24, 1

Three-dimensional photomicrograph (× 800) of stained film made from erythrocytes re-suspended in their own citrated plasma. There is a polymorphonuclear leucocyte near the centre of the field (A. Crosbie & H. Scarborough (1942), *Edinb. med. J.* **49,** 55).

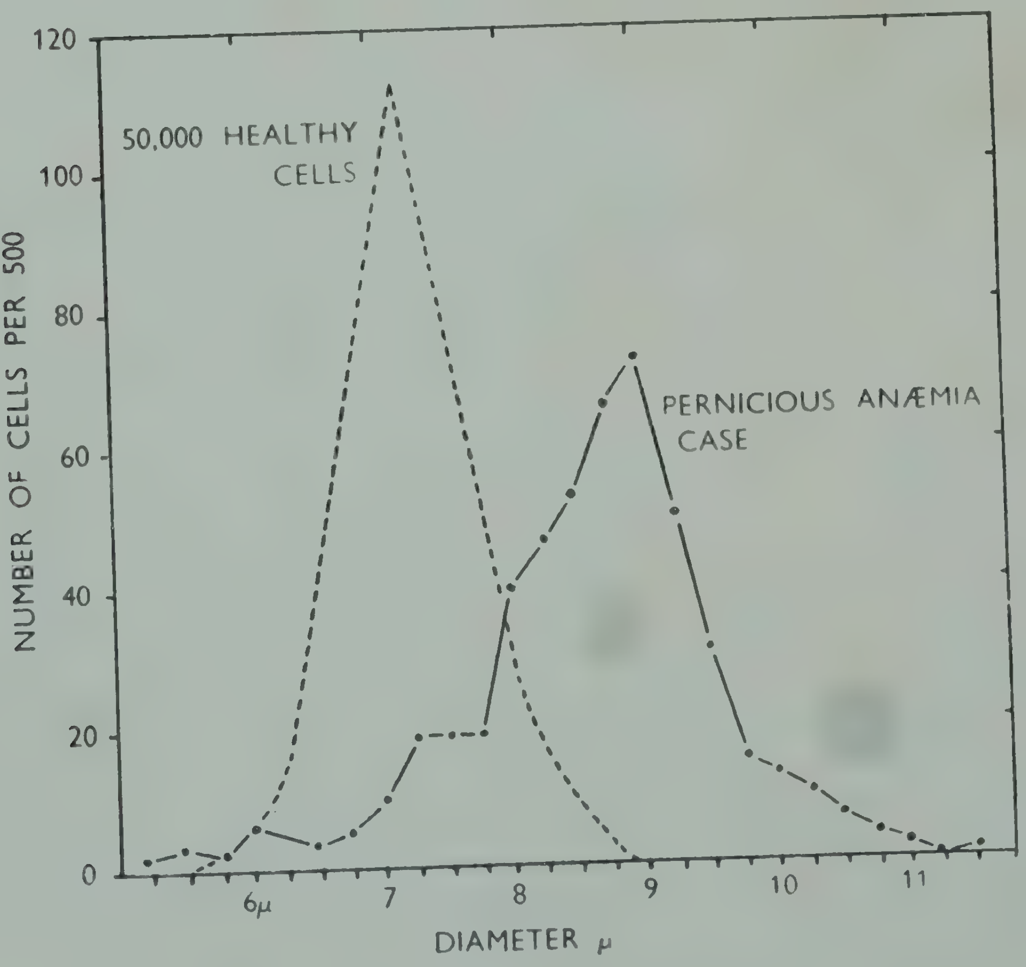


FIG. 24, 2

Frequency distribution curves of diameters of healthy red cells and of cells from a case of pernicious anæmia (C. Price-Jones (1929), *J. Path. Bact.* **32,** 487).

CHAPTER 24

THE FORMED ELEMENTS OF THE BLOOD

RED BLOOD CELLS

WHEN blood is examined microscopically without being fixed and stained, it is found to contain red blood corpuscles or *erythrocytes* and transparent bodies, white blood corpuscles or *leucocytes*, suspended in the plasma. The red cells are orange in colour when examined individually under the microscope—the red colour of blood as seen by the naked eye is due to the greater absorption of light of short wavelength (the green and blue rays) by a concentrated solution of hæmoglobin. Each red cell is a biconcave disc, the greatest thickness being near the periphery (Fig. 24, 1). The diameter of the normal red cell in plasma is about 8·8 μ with a maximum thickness of 1·9 μ. The volume of the red cells is discussed on p. 387.

A film of blood on a microscope slide is prepared for microscopic examination by staining with one of the Romanowsky stains (Leishman or Jenner) which are mixtures of methylene blue and eosin in methyl alcohol. The red cells stain a bright red colour which is more dense near the periphery where the cell is thicker. Basic stains are not taken up since the cell has no nucleus. On this account the erythrocyte can scarcely be regarded as a living cell although it normally remains in the circulation for over three months and has a measurable respiration rate. The average diameter of a red cell as measured on a stained blood film is 7·2 μ, thus the fixing and staining processes cause a shrinkage of the red cell with an average reduction in diameter from 8·8 to 7·2 μ. If the diameter of each of 500 stained red cells is measured a distribution curve of diameters can be constructed (Price-Jones' curve) from which it is evident that in health the diameters usually range from 5·5 μ to 8·5 μ (Fig. 24, 2).

If the skin is pricked through a drop of brilliant cresyl blue before a film is made and stained it is found that not more than 1 per cent. of the red cells in an adult, and some 5 to 10 per cent. in the new-born infant, show a blue staining reticulum. These *reticulocytes* are young red cells newly released from the bone marrow but the reticulum does not represent the remains of the nucleus since, unlike the nucleus, it does not take up chromatin stains. The reticulocytes may increase to as much as 50 per cent. of the red cells when blood is regenerating rapidly, as after a hæmorrhage, and an increase in the number of these young cells is, therefore, evidence of increased activity of the bone marrow.

Number of Red Cells.—It is important for diagnostic purposes to find the number of red cells present in the blood. Blood from a needle prick is drawn into a special narrow-bore pipette, is diluted with an isotonic fluid and placed on a ruled microscope slide which has a known depth between the cover glass and the ruled platform. The red cells on a given area of the platform are counted under the microscope and, since the area, the depth and the blood dilution are known, the number of red cells in 1 cu. mm of the original undiluted blood can readily be calculated.

There are normally about 5 million or 5×10^6 red cells per cu. mm, or 5×10^9 per ml., or 5×10^{12} per litre. Since the blood volume is approximately 5 litres, there are about 25×10^{12} red cells in the circulation. In the healthy adult the red cell count varies from 4·2 to $6{\cdot}4 \times 10^6$ per cu. mm, the average count in the male being $5{\cdot}5 \times 10^6$ and in the female $4{\cdot}8 \times 10^6$. However it is difficult, even with the best technique, to make a count with an error of less than $0{\cdot}25 \times 10^6$ per cu. mm. In healthy subjects there are variations of as much as 10 to 15 per cent. during the day, although these are said to be absent if the subject is in bed. The total red cell count is at its lowest value in the late afternoon. During the winter the red cell count may be increased and a change from a temperate to a tropical climate results in a fall. If the number of cells is less than 4×10^6 per cu. mm the state is described as *anæmia*. If more than $6{\cdot}5 \times 10^6$ per cu. mm are present the condition is described as *polycythæmia*. A gradual increase in the number of circulating red cells is the normal physiological response to low oxygen tension. It occurs therefore during acclimatization to life at high altitudes (Table 24, 3), and is a common finding in patients with long-standing disease of the lungs in whom the arterial oxygen saturation has been abnormally low for long periods. With rabbit bone marrow grown *in vitro* it has been found that the maximum production of red cells occurs when the atmosphere in the respirometer at 38° C contains between 10 and 20 per cent. of oxygen, that is at the tension of oxygen in atmospheric air. The production was inhibited if the percentage was greater (40) or less (5) and was completely arrested at 50 and 2·5 per cent. of oxygen.

TABLE 24, 3

This table shows the increase in red blood corpuscles which occurs when the subject is exposed to low atmospheric pressures.

Date	Altitude above sea-level		Red Cell count (cells per cu. mm)
	feet	metres	
April 10	700	215	4,480,000
May 12	4390	1350	5,240,000
May 21	8000	2460	6,040,000
May 28	10,000	3080	6,624,000
May 30	11,960	3680	6,760,000
June 1	12,400	3820	6,800,000
June 21	13,300	4090	7,525,000
June 23	15,600	4800	7,840,000
June 26	16,907	5200	7,640,000
July 27	18,203	5600	8,320,000

R. W. G. Hingston (1921), Physiological Changes at High Altitudes, *Indian J. med. Res.* **9**, 178.

The stroma, or framework, of the red cell consists of 50 per cent. protein and 10 per cent. lipids, chiefly phospholipids. Since there is more hæmoglobin in a red cell than could exist in solution in that volume of water it is probable that the hæmoglobin is adsorbed on the lipoid material. The stroma is flexible enough to allow the cells to take up an elongated shape in narrow capillaries but the normal shape is immediately regained when the cells pass into a larger vessel. The red cells are in osmotic equilibrium with the plasma which has an osmotic pressure corresponding to that of 0·85 per cent. sodium chloride. Since the surface membrane, consisting probably of lecithin and cholesterol, is freely permeable to water but not to other materials the red cell swells up if it is put into a more dilute solution of salt. The entry of water into the cell brings down its internal osmotic pressure towards that of the new environment but there is a limit to the amount of swelling which can occur without rupture of the cell. If red cells are placed in 0·42 per cent. saline half of the cells are hæmolysed, that is they burst and hæmoglobin is set free. Hæmolysis is complete in 0·35 per cent. saline. In some diseases the red cells are abnormally fragile, that is to say they hæmolyse in concentrations of sodium chloride which normal cells can resist. For some reason the resistance of the red cells to hypotonic solutions is increased by living at high altitudes. Hæmolysis can also be brought about by substances, such as ether or chloroform, which dissolve the stroma, or by acids or alkalis. By injecting red cells of one species X into an animal of species Y specific hæmolysins (p. 378) are produced in the serum of Y after an interval of time. If now the serum of Y is mixed with the cells of X, hæmolysis occurs. This reaction is so specific that it may be used for the identification of red cells as human, bovine, ovine, and so on.

Mean Corpuscular Volume.—If a red cell count is made from a specimen of blood in which coagulation has been prevented, and the blood is then centrifuged to pack the corpuscles, the relative volume of plasma and corpuscles can be obtained and the *mean corpuscular volume*, M.C.V., can be calculated thus:

$$\text{M.C.V.} = \frac{\text{Volume of packed cells in 1000 ml.}}{\text{Number of red cells in 1000 ml.}} \text{ in ml.}$$

$$\text{M.C.V.} = \frac{\text{Volume of packed cells in 1000 ml.}}{\left\{\begin{matrix}\text{Red cells in millions}\\ \text{per cu. mm}\end{matrix}\right\} \times 10^{12}} \text{ in ml.}$$

Since 1 ml. $= 10^{12}$ cu. μ,

$$\text{M.C.V.} = \frac{\text{Volume of packed cells in 1000 ml.} \times 10^{12}}{\left\{\begin{matrix}\text{Red cells in millions}\\ \text{per cu. mm}\end{matrix}\right\} 10^{12}} \text{ in cu. } \mu.$$

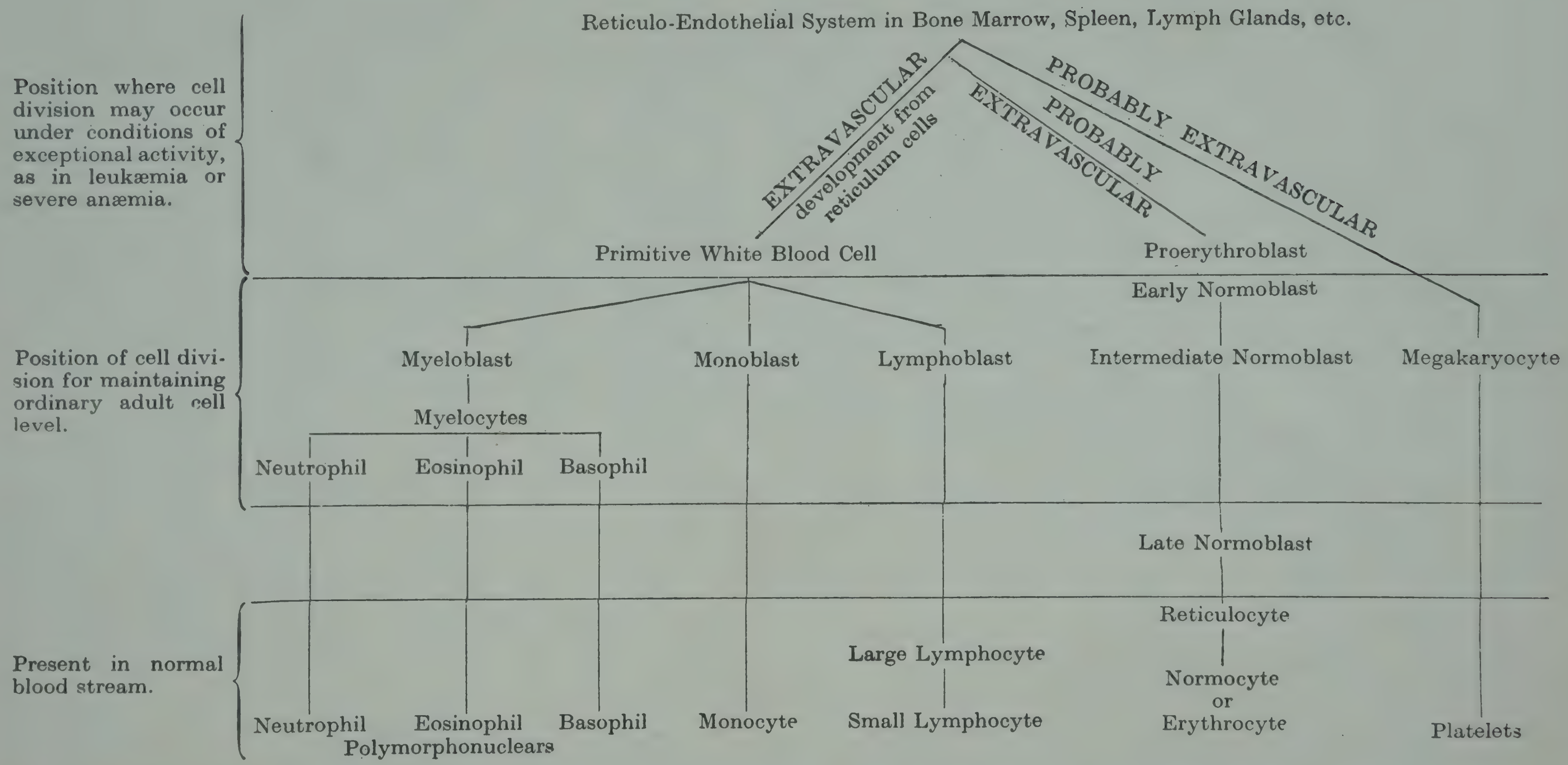
$$\text{M.C.V.} = \frac{\text{Volume of packed cells in 1000 ml.}}{\text{Red cells in millions per cu. mm}} \text{ in cu. } \mu.$$

Since the packed cell volume in 1000 ml. is about 450 ml. when 5 million red cells are present per cu. mm the average M.C.V. is $\frac{450}{5} = 90$ cu. μ.

The normal range is 78–94 cu. μ.

TABLE 24, 4

Origin of Blood Cells (Polyphyletic Theory)


Reticulo-Endothelial System in Bone Marrow, Spleen, Lymph Glands, etc.
EXTRAVASCULAR
development from reticulum cells
PROBABLY EXTRAVASCULAR
PROBABLY EXTRAVASCULAR
Position where cell division may occur under conditions of exceptional activity, as in leukæmia or severe anæmia.
Primitive White Blood Cell
Proerythroblast
Position of cell division for maintaining ordinary adult cell level.
Early Normoblast
Myeloblast
Monoblast
Lymphoblast
Intermediate Normoblast
Megakaryocyte
Myelocytes
Neutrophil
Eosinophil
Basophil
Late Normoblast
Present in normal blood stream.
Reticulocyte
Large Lymphocyte
Normocyte or Erythrocyte
Neutrophil
Eosinophil
Basophil
Polymorphonuclears
Monocyte
Small Lymphocyte
Platelets

Note: The early stages in the development of the blood cells are difficult to recognize and the above table represents only one view of their origin.

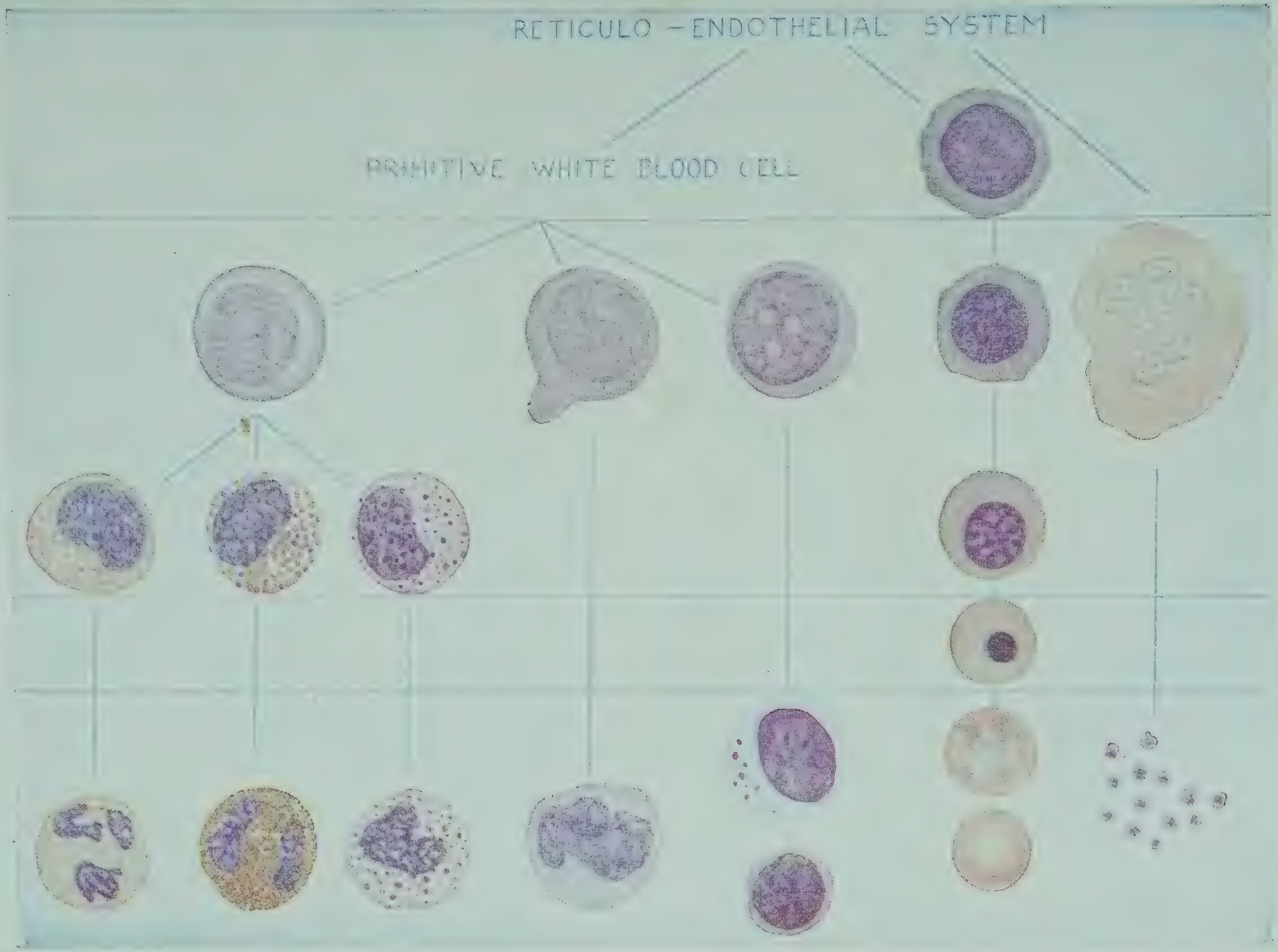

RETICULO - ENDOTHELIAL SYSTEM
PRIMITIVE WHITE BLOOD CELL

FIG. 24, 5

Scheme adopted by Whitby and Britton. Note the absence of megaloblasts in the normal erythropoietic series. This cell is found only in the embryo or in the blood in pernicious anæmia. The names of the cells in this coloured plate are given in Table 24, 4, on the facing page. (From L. E. H. Whitby and C. J. C. Britton (1950), *Disorders of the Blood*, 6th edition, p. 6. London: Churchill.)

[*Facing page* 388

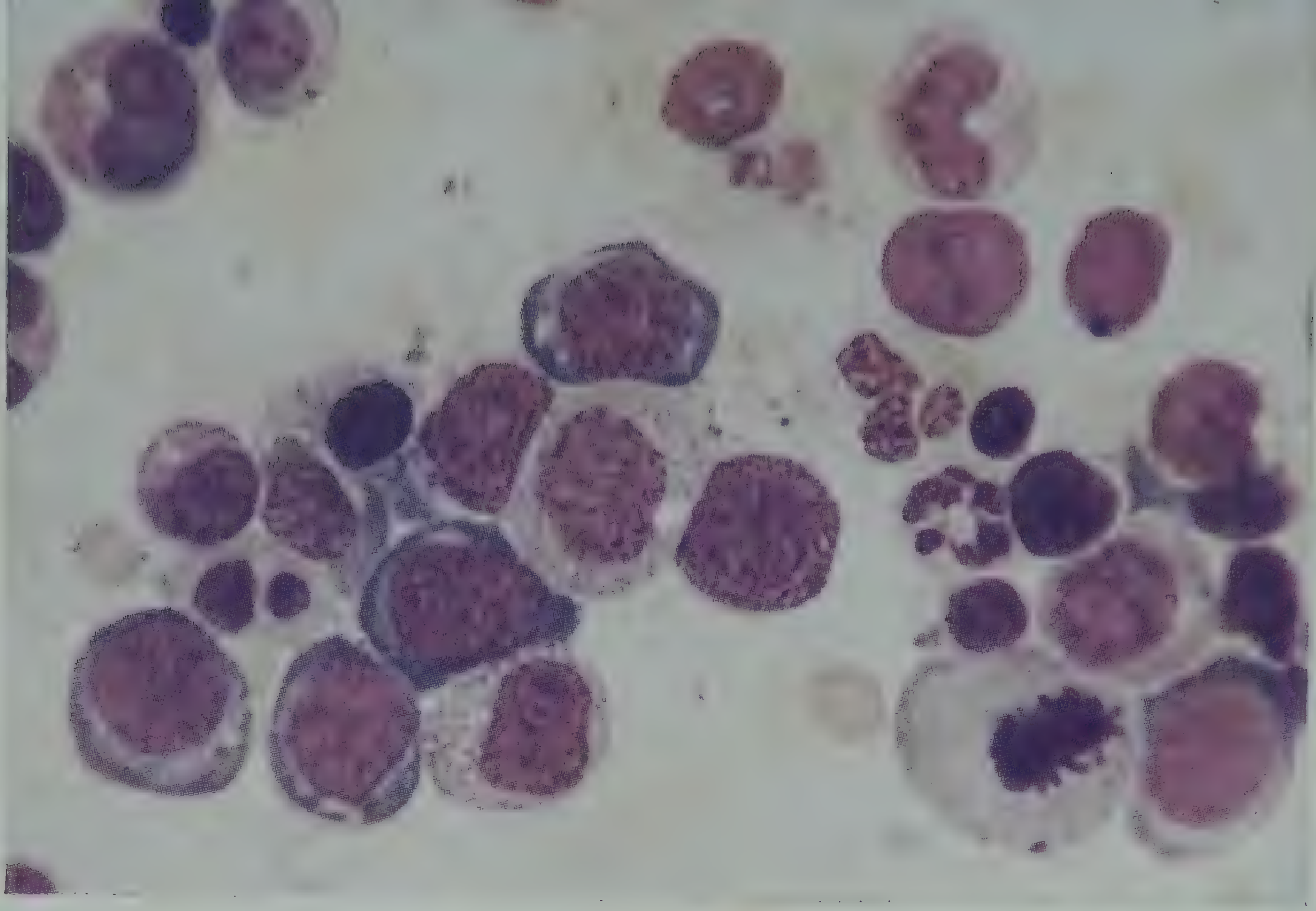


FIG. 24, 6A

Marrow smear showing megaloblastic erythropoiesis in an untreated patient with pernicious anæmia. The large early megaloblasts with basophilic cytoplasm and finely stippled nuclei are easily recognized; only one or two normoblasts can be seen. Note also the large metamyelocytes and the polymorph showing hypersegmentation of its nucleus. (Stained Jenner-Giemsa; ×1000.)

(*Reproduced from* The Glaxo Volume *by courtesy of J. N. Marshall Chalmers and Glaxo Laboratories Ltd.*)

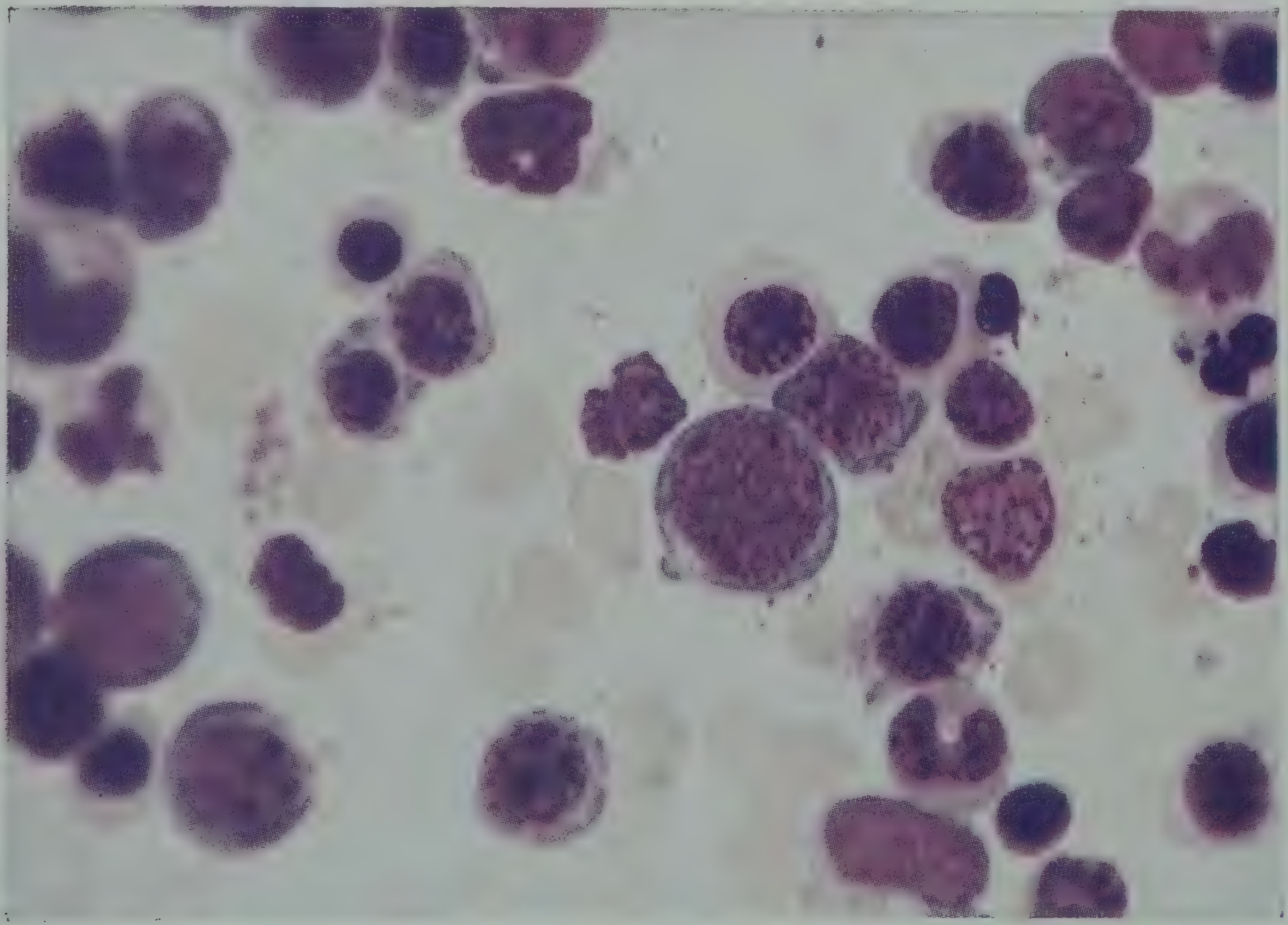


FIG. 24, 6B

Marrow smear showing normoblastic erythropoiesis in a patient with pernicious anæmia 48 hours after intramuscular injection of 50 μg. of vitamin B_{12} ('Cytamen'). In contrast to illustration A, there is here normoblastic hyperplasia, megalo blasts being no longer demonstrable. The cell in the centre is a proerythroblast, a physiological red cell precursor. (Stained Jenner-Giemsa; ×1000.)

(*Reproduced from* The Glaxo Volume *by courtesy of J. N. Marshall Chalmers and Glaxo Laboratories Ltd.*)

Sedimentation Rate.—When blood to which an anticoagulant has been added is allowed to stand the red blood corpuscles form aggregates or *rouleaux* in which the cells are arranged face to face like a pile of coins and the aggregates gradually settle down (sediment) to leave a clear zone of plasma above. This phenomenon is related to the fact that each red cell has a negative electric charge; anything that increases the positive charge in the plasma favours rouleaux formation and the larger the rouleaux the more rapidly do they sediment. The rate of sedimentation, as indicated by the length of the column of clear plasma at the end of an hour, is increased in acute and chronic infections, returning gradually to normal as the infective process declines. The sedimentation rate is, therefore, of value in diagnosis and also in assessing the results of treatment but the test is unspecific and the results must be interpreted with caution. The suspension stability of the red cells as expressed by the sedimentation rate is governed by quantitative and qualitative changes in the proteins of the plasma; thus an increase in the concentration of fibrinogen and globulin accelerates the rate and a high concentration of albumin retards it. The sedimentation rate tends to vary according to the red cell count, the more anæmic the blood the more rapid being the sedimentation rate.

Origin and Development of the Blood Cells

In the first two months of intrauterine life the blood cells are formed in the yolk sac, body stalk and placenta. The earliest red cells are shed from the endothelium of the blood vessels into the lumen of the vessel. When hæmoglobin develops in them these nucleated cells are known as *megaloblasts*. These cells multiply in the blood stream to give smaller nucleated hæmoglobinized red cells which later lose their nuclei and become erythrocytes or normocytes. The liver, together with the spleen and thymus, is the chief site of production from the second to the fifth month. After this blood cell formation begins in the bone marrow which soon takes over all hæmopoietic activity, so that at birth all the red cells and the granular leucocytes are formed in the bone marrow. The liver and spleen remain, however, potential hæmopoietic organs and during childhood can produce red cells in response to a greatly increased demand (*extramedullary hœmopoiesis*). White cells are not produced in great numbers till the fœtus is about five months old. These cells are formed extravascularly in the mesoderm and migrate into the blood vessels.

The origin of the cells of the blood after birth is still a matter of controversy. The generally accepted view given here is summarized in the colour diagram of Whitby and Britton (Fig. 24, 5), (Table 24, 4). From the fixed cells of the reticulo-endothelial system all the cells of the blood are eventually derived by division and maturation. It seems best to assume that at the first division of the reticulum cells the daughter cells are destined to become either red cells or white cells. The supply of blood cells is probably kept up by divisions of these daughter cells and not by division of undifferentiated reticulum cells. Although there is still some doubt as to whether the precursors of the red cells are formed intravascularly or extravascularly, the white series seem undoubtedly to be formed from precursor cells outside the vessels which they later enter. Megakaryocytes are probably formed extravascularly by fusion of several reticulum cells.

Erythropoiesis in the Bone Marrow.—Until the age of four years all the bones contain actively hæmopoietic red marrow but thereafter the marrow of long bones becomes more and more yellow, due to the gradual replacement of the active marrow by fatty tissue. From the age of twenty onwards only the flat bones, vertebræ, sternum, ribs and skull contain appreciable amounts of red marrow, although in the adult femur and other long bones red marrow can be seen on microscopic examination. After repeated hæmorrhages the volume of red marrow increases considerably in the flat bones and the femur and humerus may also show an increase in red marrow which extends from the small foci of hæmopoietic tissue in the ends of these bones.

Samples of human bone marrow can be obtained by pushing a short wide needle through the outer table of the sternum into the marrow cavity (*sternal puncture*). A small amount of marrow juice is withdrawn into a syringe containing oxalate and smears are made and stained (Fig. 24, 6). These show large numbers of red cells without nuclei and a smaller number of nucleated cells. Since the volume of bone marrow devoted to erythropoiesis is much smaller than that devoted to leucopoiesis the majority of the nucleated cells in a sternal puncture smear are immature white cells. Examination of material obtained by sternal puncture is an important diagnostic procedure.

The successive early stages of the development of the immature cells (Figs. 24, 5, and 24, 6) in the bone marrow are difficult to differentiate. The megaloblast, the characteristic immature cell in the developing embryo, is not normally seen much after birth and in extrauterine life erythropoiesis is described as normoblastic. In its early stages the normoblast has a large nucleus with an open network of chromatin and basophilic cytoplasm. As the cell becomes older it decreases in size, the nucleus becomes more dense and hæmoglobin appears in the cytoplasm. Finally the nucleus either fragments and dissolves or is extruded from the cell, which is then at the reticulocyte stage. Large numbers of reticulocytes are found in normal bone marrow but only a few in the circulating blood. The primitive hæmopoietic cells have a relatively high concentration (over 5 per cent.) of ribonucleic acid; as the cells mature and mitosis ceases the ribonucleic acid content falls to 0·5 per cent. At this stage hæmoglobinization begins in the erythroblasts, the hæmoglobin content rising rapidly at first and later more slowly.

Life of the Red Cell.—When the red cells in the circulation become old they fragment, and the fragments are taken up by the reticulo-endothelial cells of the liver, spleen and bone marrow. There is evidence, too, that the tissues contain lytic agents with which the red cells come into contact during their circulation through capillaries. The absence of any morphological difference between old and new red cells makes the estimation of their survival time rather difficult but recently the application of several independent methods has led to the conclusion that the maximum period of survival is about 120 days.

The first method, the method of differential agglutination, depends on the agglutination of red cells by serum agglutinins, as described on p. 380. A transfusion of red cells of blood group O is given to a person of blood group A. Blood samples are taken at intervals and treated with anti-A serum which lyses all the A cells; the O cells remaining are counted.

The number of O cells gradually declines, the rate of destruction of transfused red cells being about 0·8 per cent. of the initial amount per day. In 60 days after the transfusion 60 × 0·8 per cent. of the group O cells are destroyed. In other words, about half of the transfused cells live 60 days and all disappear in about 120 days. The average life of a transfused red cell is thus 60 days. The variation in the period of survival of the cells is due to the fact that at the time of transfusion the sample of group O red cells contains cells of all ages; the new ones just released from the bone marrow presumably survive a long time while the old ones just due to be scrapped are probably destroyed very soon. If a normal person treats his own cells as he treats transfused cells the average age of his red cells must be about 60 days and the total life span of an individual red cell in the circulation must be about 120 days (Fig. 24, 7).

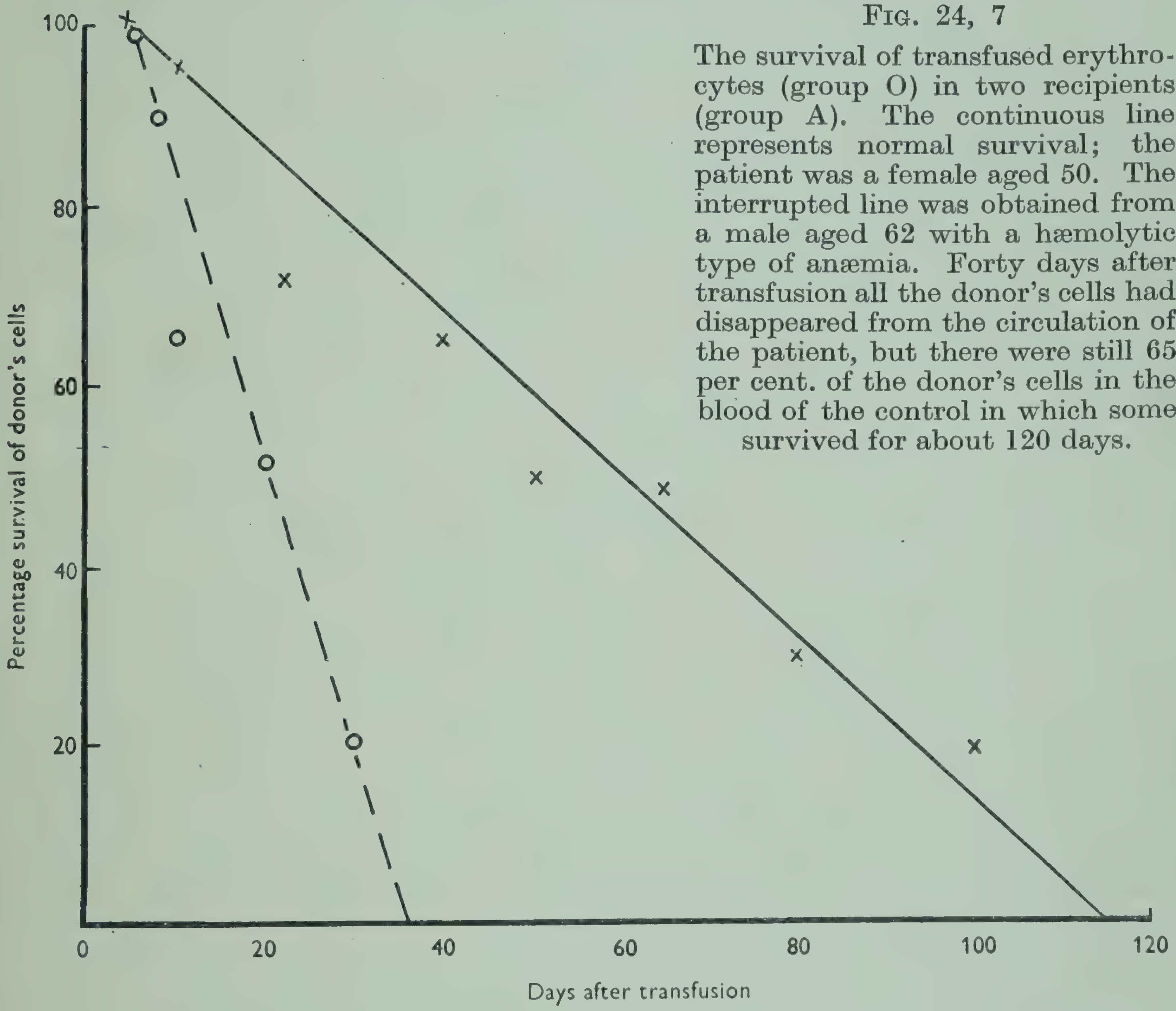


FIG. 24, 7

The survival of transfused erythrocytes (group O) in two recipients (group A). The continuous line represents normal survival; the patient was a female aged 50. The interrupted line was obtained from a male aged 62 with a hæmolytic type of anæmia. Forty days after transfusion all the donor's cells had disappeared from the circulation of the patient, but there were still 65 per cent. of the donor's cells in the blood of the control in which some survived for about 120 days.

If glycine labelled with ^{15}N is given to a human being the isotope is detectable in the porphyrin ring (p. 411) of the hæmoglobin in the red cells on the fourth day, and the concentration rises steadily over a period of 25 days and then remains constant for a further 50 days. The isotope concentration then drops slowly to zero in the next 150 days. This is a rather unusual finding, as the concentration of a labelled compound usually rises in the tissues as synthesis proceeds and then declines as the labelled molecules are metabolized. To explain the 50-day period during which the red cell content of ^{15}N is constant it is necessary to assume that the porphyrin part of the hæmoglobin of the new red cell is excluded from the metabolic reactions of the body and returns to these only when the red cell is broken up. From these results it has been calculated that the total life span of a human red cell is 127 days. This method shows incidentally that the porphyrin group, in contrast to the iron, is not retained to make new red cells but is excreted as bile pigments in the bile.

Nutritional Factors in Hæmopoiesis.

In health the number of red cells and the amount of hæmoglobin in the blood remain remarkably constant since the destruction of old red cells is balanced by the formation of new ones at the rate of approximately 0·8 per cent. per day. This process is an extremely complicated one because of the variety of substances required for hæmopoiesis.

Although the body retains its iron most carefully, a diet which cannot make good the small amount lost eventually produces anæmia. In a mild iron deficiency the red cells may not be much reduced in number but each cell contains less than its normal complement of hæmoglobin, that is, the mean corpuscular hæmoglobin concentration is reduced (p. 407). The healthy adult absorbs daily about 1–3 mg. of iron, equivalent to some 10 per cent. of the iron content of a normal diet. The amount of iron absorbed can, however, be regulated according to the needs of the body, for when these are large iron is more readily absorbed from the alimentary tract, while absorption is relatively less efficient when the iron reserves of the body are good (see p. 98). By feeding radioactive iron ^{59}Fe it has been found that the peak absorption of iron occurs four to eight hours after a meal, that is, at a time when the food materials are in the small intestine. After absorption from the small intestine in the ferrous state the iron becomes linked with a protein *apoferritin* in the intestinal cell to form *ferritin* (Fig. 9, 6). At the point of the cell in contact with the blood iron is transferred from ferritin to a plasma globulin *transferrin* and distributed to the tissues where it is stored as ferritin and to the hæmopoietic organs where it is utilized in the formation of hæmoglobin. Iron is present in plasma in the ferric state to the extent of 100 to 120 μg. per 100 ml. but there is considerable diurnal variation in concentration, the lowest values being found in the late evening and the highest in the morning. The transport of iron is a function of transferrin, one of the β-globulins of the plasma protein, 1 g. of β-globulin being capable of binding 1·25 mg. of iron. Normally, however, only about 35 per cent. of the iron-binding capacity is taken up, the remainder representing a margin of safety for the transport of additional iron should the need arise.

The reserves of iron of a healthy man amount to some 600 mg., sufficient to replace about 25 per cent. of the circulating hæmoglobin. This iron is mainly in the liver in the form of ferritin. When, however, the capacity to form ferritin is fully taken up, the additional iron is deposited both in the liver and in other tissues as *hæmosiderin* (p. 98). The iron of both ferritin and hæmosiderin is available for hæmoglobin synthesis.

The absorption of iron from the alimentary tract is defective when there is achlorhydria because then the iron is not liberated from organic complexes present in food. Diarrhœa, obstructive jaundice, impaired fat absorption, and large amounts of phytic acid in the diet also interfere

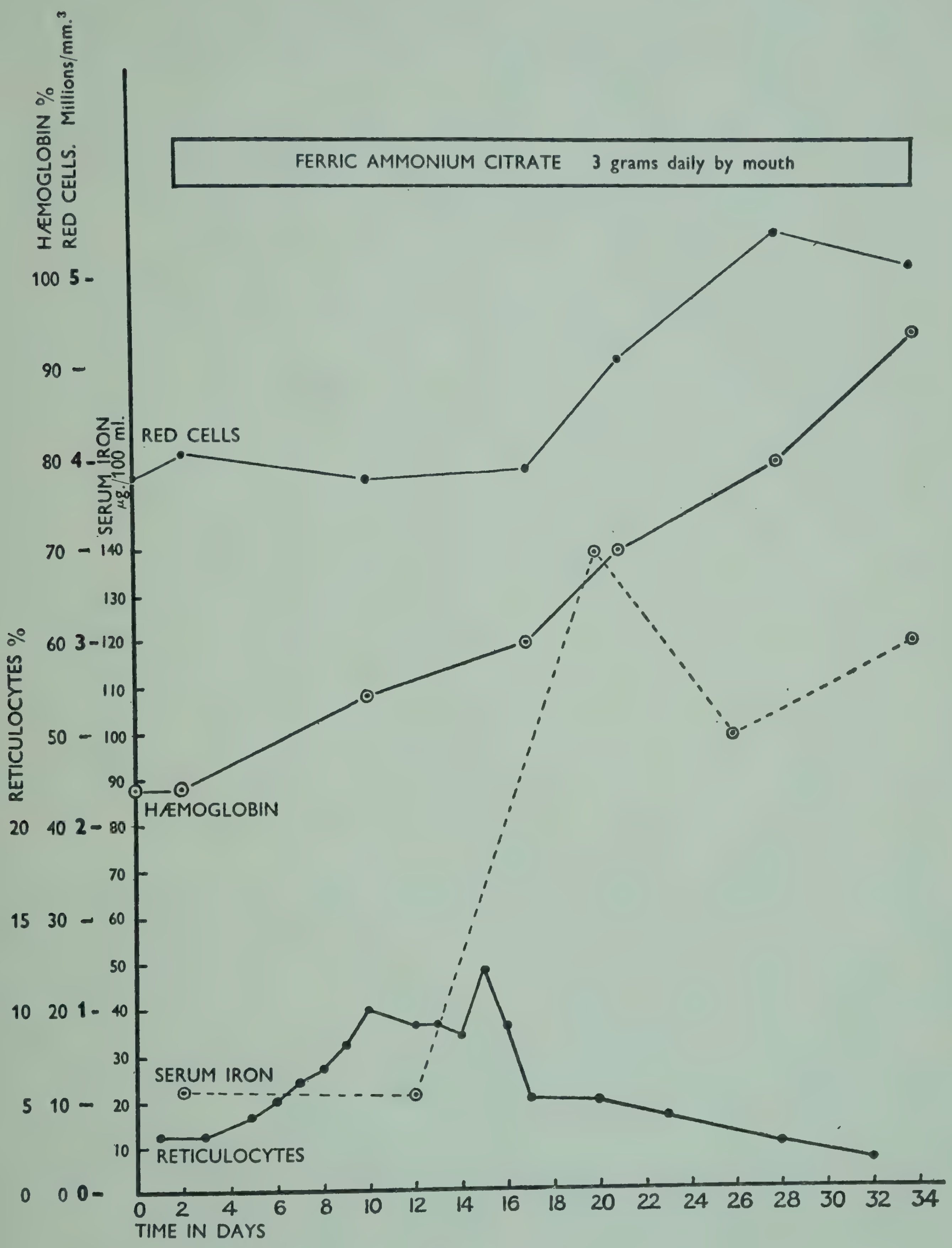


FIG. 24, 8

The effect of oral administration of iron on the blood of a female patient aged 74 with an iron-deficiency anæmia. The serum iron, initially very low, rises to normal levels during the reticulocytosis which is maximal some 8 to 10 days after beginning treatment with iron.

with absorption. These conditions are liable to produce anæmia or to interfere with the treatment of anæmia produced by long-continued bleeding. In these circumstances administration of iron by the intravenous route may be effective. Anæmia is commonly found in chronic infections and in other prolonged illnesses; the defect probably is a faulty utilization of iron by the bone marrow. Little or no iron is eliminated in the urine and only very small quantities are excreted into the intestine but the sweat contains appreciable amounts (p. 99).

In a dog made anæmic by repeated bleedings hæmoglobin is very quickly regenerated. Radioactive iron can be detected in the circulating red cells four hours after feeding and the radioactive iron which is absorbed is all converted to hæmoglobin within four to seven days.

Whipple has made much use of the dog to study the effects of various foods on blood formation. By repeated daily bleedings of appropriate amounts the hæmoglobin can be kept nearly constant within the range 40 to 50 per cent. of normal. The amount which has to be withdrawn to bring down the blood to this level of anæmia is a measure of the amount of blood produced during the experimental period. In 1925 he found that the maximum regeneration of hæmoglobin and red cells occurred when ox liver was fed. This suggested that liver contained some important precursor of hæmoglobin or some factor which accelerated hæmopoiesis. After these experiments on dogs and reports of the successful treatment of a few patients, two American physicians, Minot and Murphy, in 1926 gave liver by mouth to a number of patients suffering from pernicious anæmia. The dramatic improvement that followed was actually the result of a fortunate coincidence, because we know now that the liver contains two quite distinct factors of different solubilities, one of which causes the regeneration of blood in dogs made anæmic by bleeding and the other of which cures pernicious anæmia in man.

In pernicious anæmia the development of the red cells (erythropoiesis) is disturbed. Instead of maturing into normoblasts and normocytes the pro-erythroblasts give rise to immature cells called megaloblasts which later lose their nuclei and appear as large red cells, called macrocytes. These cells enter the peripheral circulation which therefore contains cells on the average bigger than normal blood cells (see Price-Jones' curve, Fig. 24, 3). If the patient is not treated the number of red cells and the amount of hæmoglobin in the circulating blood decline and the patient ultimately dies unless he is given liver or liver extract. This permits return to the normal course of hæmopoiesis and large numbers of normocytes reach the peripheral blood. Liver must, therefore, contain an erythrocyte-maturing factor (or anti-pernicious anæmia principle).

A number of experiments carried out on patients by Castle suggested that the hæmopoietic factor is formed by the interaction of an *extrinsic factor* which is obtained from the food with an *intrinsic factor* secreted by normal gastric mucosa. The anti-pernicious anæmia principle formed is absorbed from the intestine and stored in the liver. In pernicious anæmia the secretory functions of the gastric mucosa are greatly impaired and neither hydrochloric acid nor intrinsic factor are secreted. Administration of normal gastric juice alone, or beef alone, does not cure the condition, but if beef is incubated *in vitro* with normal gastric juice the resulting product can cure pernicious anæmia. Moreover, normal human gastric juice, though inactive when fed to patients with pernicious anæmia with empty stomachs, can produce full remission when administered to patients who have been given beef, without prior digestion, immediately beforehand. Further experiments on these lines showed that the intrinsic factor is not pepsin, hydrochloric acid, rennin, or lipase; it is present only in gastric juice and not in duodenal juice. The extrinsic factor was similarly demonstrated in beef muscle, liver, eggs, and substances which are good sources of the vitamin B complex. Investigation of the chemical nature of the factors is greatly hindered by the fact that true pernicious anæmia occurs only in human beings and has not yet been produced experimentally in animals. All estimations of the potency of extracts of liver and other tissues have, therefore, to be made on patients suffering from pernicious anæmia. As early as six hours after the injection of a potent extract to such a patient cytological changes are observed in the bone marrow, the large number of primitive megaloblasts being replaced by more mature cells. Three to four days later there are few megaloblasts in the bone marrow, the predominant erythroblastic cell being the normoblast. In the circulating blood the number of reticulocytes gradually increases to reach a maximum on the third or fourth day after injection, after which it declines (Fig. 24, 9). Soon after 1926 liver extracts became available with a potency one hundred times or more that of liver itself and later fractions fully potent in minute doses were prepared.

In 1949, cyanocobalamin (vitamin B_{12}) (p. 132) was isolated. It was soon found that complete remission of pernicious anæmia regularly followed the intramuscular injection of this material, though not its administration by mouth unless normal gastric juice was given at the same time. Doses of 0·5 μg. of vitamin B_{12} labelled with ^{60}Co given by mouth can be traced to the fæces by their radioactivity. In healthy people about one-third of the test dose is recovered in the fæces but in persons with pernicious anæmia 90 per cent. of the test dose is recovered. Pernicious anæmia patients given neutralized normal human gastric juice, or watery extracts of gastric mucosa, as sources of intrinsic factor lost a substantially smaller quantity of the vitamin in the fæces. The simplest, but not necessarily the soundest, explanation of these facts is that cyanocobalamin is equivalent both to Castle's extrinsic factor and to the anti-pernicious anæmia principle; gastric intrinsic factor, present in normal gastric juice, but not in that from persons with pernicious anæmia, is necessary for its absorption from the alimentary tract. In the pig intrinsic factor is secreted by the mucosa of the pyloric part of the stomach and of the duodenum. Intrinsic factor activity has been found in a mucoprotein prepared from gastric glands.

Apart from iron and the anti-pernicious anæmia factor a number of other factors is involved in hæmopoiesis. One of the most important of these is protein. Whipple's studies on anæmic dogs have demonstrated that blood regeneration is accelerated when adequate protein is supplied in the diet. The value of copper in the regeneration of blood has been demonstrated in animals and also in children but anæmia due solely to a dietary lack of copper is unknown in man. The anæmia which occurs in myxœdema (Chapter 51) is cured by giving thyroid gland and the anæmia not infrequently found in scurvy (p. 140) is corrected by the administration of ascorbic acid. How these substances act in promoting hæmopoiesis is not at present known. An ordinary, good, mixed, diet contains sufficient protein, vitamins, minerals and extrinsic factor for normal hæmopoiesis.

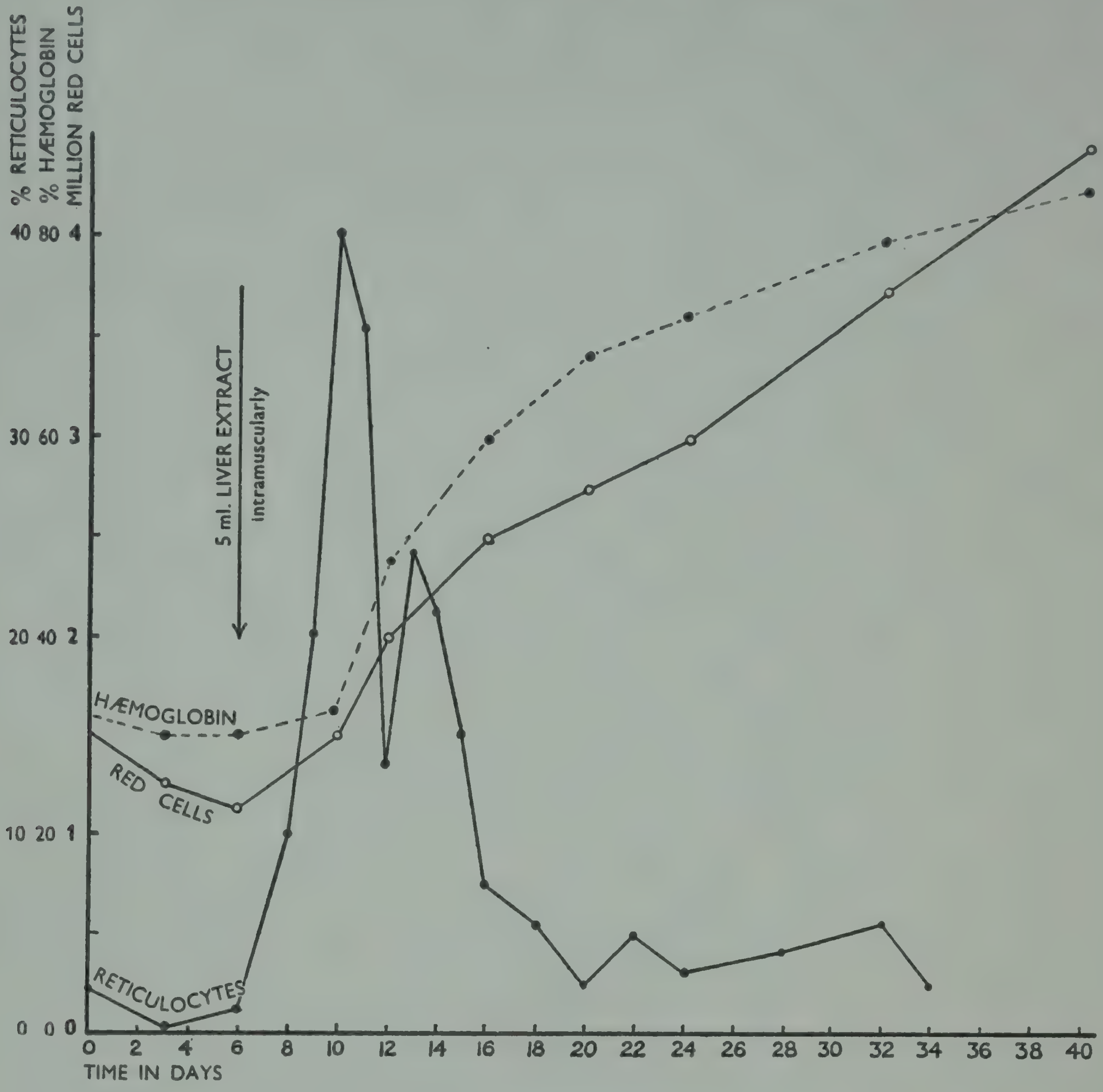


Fig. 24, 9

Illustrating the response to the administration of liver extract in a male aged fifty-seven with Addisonian pernicious anæmia. The maximum reticulocytosis occurred on the fourth day after the injection and from this time there was a steady increase in both hæmoglobin and circulating red cells.
(100 per cent. hæmoglobin ≡ 14·8 g. per 100 ml. blood.)

LEUCOCYTES

In addition to the red cells there are in the circulating blood colourless cells called *white cells* or *leucocytes* which, unlike the red cells, contain nuclei. Leucocytes consume oxygen and produce lactic acid from glucose. They have been shown to contain amongst other substances ascorbic acid, glutathione, histamine and glycogen and possess activity indicative of the presence of amylase, lipase, catalase, nucleotidase, phosphatase, β-glucuronidase and proteolytic enzymes.

The three varieties of white cells—the *granular* or *granulocyte series* (so called because of their numerous cytoplasmic granules), the *lymphocytes* and the *monocytes*—are all derived from the reticulo-endothelial system (Fig. 24, 4). From the time of birth onwards all the granular cells are formed in the bone marrow and they are, therefore, often termed the *myeloid* series. Although a few lymphocytes are formed in the marrow, the majority are produced in the lymphatic tissue of the body. The site of production of the monocytes is not known with certainty.

The myeloid cells are formed extravascularly from the non-granular myeloblast which gives rise to three varieties of myelocytes, *neutrophil*, *eosinophil* and *basophil*, all of which contain granules. When these cells mature they become the *neutrophil*, *eosinophil* and *basophil polymorphonuclear* cells of the blood stream. The polymorphonuclear cells are so called because their lobed nuclei vary greatly in shape, the number of lobes increasing with the age of the cell up to four or more. The neutrophil polymorphs, about 10 to 12 μ in diameter, are actively amœboid and phagocytic when examined fresh on a warm stage although they are probably inactive in the circulating blood. The small cytoplasmic granules stain faintly pink with Romanowsky stains and are, therefore, not quite strictly neutrophil. The eosinophil cell, about 10 to 12 μ in diameter, the nucleus of which is usually bi-lobed, contains much coarser, red-staining, granules scattered through the cell. The eosinophils may show amœboid movements but they are not actively phagocytic. The basophil cell is somewhat smaller, 8 to 10 μ, with large cytoplasmic granules which stain a deep blue; the granules of these cells, sometimes called 'mast' cells, contain heparin and, almost certainly, histamine.

The granules of the myeloblast series stain brown with a mixture of alpha-naphthol and hydrogen peroxide. This *oxidase reaction* is, however, not well marked in the immature cells of the series although it is often useful in determining to which series immature cells belong. Monocytes, on the other hand, have few oxidase granules while the lymphocyte series show none. Leucocytes of the granular series contain an alkaline phosphatase whose activity may normally be expressed as from 25 to 60 units. When infection is present in the body, even although there is no leucocytosis, the phosphatase activity increases some ten- to twenty-fold. In the disease *leukœmia*, however, which is characterized by greatly increased numbers of leucocytes in the blood and tissues, the phosphatase activity is reduced below the normal.

It has been possible by special tissue-culture methods to keep bone marrow alive and to observe in it multiplication and maturation of the granulocyte series of cells for at least six weeks. When this method was applied to human blood it was found that neutrophils survived about 60 hours on the average. This is the best estimate we have of the life of these cells in the peripheral blood. Under the same conditions eosinophils live 8 to 12 days and basophils 12 to 15 days. The old white cells are taken up by the reticulo-endothelial cells.

It is becoming apparent that the lungs may influence the numbers of leucocytes in the circulating blood. Thus in man after intravenous injection of 0·2 mg. of adrenaline higher concentrations of white cells were found in blood taken from the femoral artery than in samples simultaneously withdrawn from the pulmonary artery (by catheter). On the other hand, there is evidence both in man and in rabbits that large numbers of leucocytes can be lost from the circulating blood as it passes through the lungs.

The monocyte, 16 to 22 μ in diameter, the largest cell in blood, is recognized by its oval or horseshoe-shaped nucleus. The cytoplasm of the monocyte is pale-staining and frequently contains vacuoles and a few small reddish granules. Its staining reactions suggest that the monocyte is more closely related to the myeloid series than to the lymphocytes. The monocytes probably originate mainly in the spleen but this is by no means proved. Many hæmatologists regard these cells as histiocytes which have migrated into the blood stream (p. 402), implying that they are 'free' reticulo-endothelial cells. Monocytes show slow amœboid movements and are phagocytic.

The lymphocytes arise by proliferation of the reticulum cells in the germ centres of the lymph nodes. The first cell of the developmental series is a large cell, the *lymphoblast*, which by dividing repeatedly gives rise to the *large lymphocyte*, 12 to 15 μ in diameter, with abundant cytoplasm and round or indented nucleus. These cells appear in the blood stream. By a process of maturation the large lymphocytes become *small lymphocytes* about 7 μ in diameter, with deeply staining nuclei and very scanty cytoplasm. Neither the large nor the small lymphocytes are motile in ordinary preparations; although lymphocytes can be shown to have amœboid movement they are not phagocytic. Cells intermediate in size and appearance between the large and small lymphocytes can always be seen. Large lymphocytes stained with Leishman's stain usually show one or two reddish coarse granules and small lymphocytes occasionally possess similar granules.

From lymph nodes the majority of the lymphocytes pass by way of the lymphatic vessels to the thoracic duct and so gain entrance to the blood stream. When the thoracic duct is separated from the veins the fall in the total number of lymphocytes in the blood is accounted for by the number collected in the same time from the free end of the thoracic duct. Since the production rate is equal to the disappearance rate the number of lymphocytes in the circulation is relatively constant. If the lymph from the thoracic duct is collected over twenty-four hours it is found that the total number of lymphocytes contained in it is about three to five times that present in the circulating blood at one time. The stay of a lymphocyte in the peripheral blood cannot, therefore, be more than a few hours. However, lymphocytes have been observed through transparent chambers inserted into the ear of rabbits and in some cases they have been active and motile in the tissues for over a week.

There is still considerable ignorance about the fate of the lymphocytes. Many are normally lost in the saliva and in the secretions of the gastro-intestinal tract though lymphocytes disappear rapidly from the blood of animals even when the stomach and intestine have been removed. Others appear to return to the blood stream, in which they reach the germinal centres of the lymphoid tissue to be disintegrated by the numerous macrophages found there. The macrophages retain the nuclear material but the cytoplasm of the lymphocytes is liberated into the general circulation. This extrahepatic source of protein is especially rich in γ-globulin and, therefore, in antibodies. There is strong evidence that the lymphocytes in lymph nodes can form antibodies and that they retain this power when transferred to a new host. Thus, after bacteria had been injected into the hind limb of a rabbit, the popliteal lymph node was excised, the lymphocytes isolated, washed and injected intravenously into a normal rabbit.

The blood of this rabbit was then found to contain antibody specific for the bacteria injected into the hind limb of the first.

Enumeration of the White Cells.—The method is similar in principle to that for red cells (p. 385). The diluting fluid contains acetic acid and gentian violet. The acid lyses the red cells, which disappear, but the nuclei of the white cells are not destroyed and, because they are stained and refractile, are easily seen and counted. A *differential count* is needed to estimate the number of each variety of white cell. A blood film is made and stained; at least 500 cells are counted and each is entered in its proper category. The proportion of each cell being known, the number per cu. mm can be calculated from the total white cell count (Table 24, 10).

TABLE 24, 10

A list of the White Cells found in normal adult blood. The numbers occurring in the circulating blood usually fall within the ranges shown

	Number per cu. mm	Percentage of total white cells
All leucocytes . . .	5000–9000	100
Neutrophil polymorphonuclears	3000–6000	60–70
Lymphocytes . . .	1500–2700	20–30
Monocytes	100–700	2–8
Eosinophil polymorphonuclears	100–400	2–4
Basophil polymorphonuclears .	25–200	0·5–2

The number of leucocytes in the circulating blood is very variable but is usually between 5000 and 9000 per cu. mm. An increase in the number of leucocytes above 10,000 per cu. mm is called a *leucocytosis* and a fall below 4000 a *leucopenia*. Violent exercise, fever and hæmorrhage are followed by a leucocytosis and some increase occurs a few hours after a protein meal. Damage to, or destruction of, tissue as in accidents, fractures, surgical operations and infections leads to considerable and often prolonged increase in circulating polymorphonuclear cells, the number of which may rise markedly within a few hours. How this rapid and sustained increase is brought about is not known, but it seems likely that the stimulus to increased production is a humoral one, possibly associated with nucleic acid derivatives produced as the result of cellular destruction. Leucopenia occurs in starvation or debilitating conditions. While acute infections characteristically increase the production of neutrophil polymorphs, chronic infections usually increase the lymphocyte count.

When an increase in the numbers of circulating neutrophil cells (polymorphonuclear leucocytosis) occurs, for example as the result of an infection, examination of a stained blood film shows a higher than normal proportion of neutrophil cells with nuclei containing one, two or three lobes instead of the four or five of the mature polymorphonuclear corpuscles. Since the cells with the morphologically more simple nuclei are less mature forms of neutrophil leucocyte and because their development is visualized as proceeding from left to right this phenomenon is sometimes referred to as a 'shift to the left'. It may be important in clinical medicine as indicating that the bone marrow is reacting even when there is no leucocytosis.

In certain diseases of the skin, in parasitic infestations and in the allergic group of disorders which includes asthma and hay fever an increase in the eosinophil count (*eosinophilia*) is often found. The concentration of circulating eosinophils varies by as much as 50 per cent. during the day. The lowest counts with the least variation are found during the early afternoon. The concentration of circulating eosinophils is depressed by cortisone and adrenocorticotrophic hormone (Chapter 51) so that the fall in the eosinophil count which is maximal about 4 hours after the injection can be used as a test of the activity of these materials.

THE PLATELETS

The megakaryocytes of the bone marrow, believed to be the main source of the blood platelets, are large cells some 20 to 40 μ in diameter derived from reticulo-endothelial cells. Since the original nucleus divides repeatedly while the cytoplasm remains undivided, each cell comes to have a large number of nuclei in its cytoplasm. These cells, some of which are phagocytic and many of which show amœboid movements, rarely appear in the peripheral blood (Fig. 24, 5). When grown in tissue culture they slowly rotate.

Megakaryocytes throw out pseudopodia which, passing through the walls of sinusoids in the marrow, may be nipped off to become platelets. However, in very hyperplastic marrows it appears that platelets are formed within the body of the megakaryocytes and released by rupture of the cellular membrane. It is possible that giant cells in the lungs also produce platelets, because blood coming from the lungs contains a larger number of platelets than venous blood going to the lungs. On the other hand, there is a reduction in the number of platelets as the blood traverses the systemic circulation.

The platelets, usually about 2 or 3 μ in diameter and the smallest of the formed elements in blood, are cytoplasmic fragments containing granules but no nucleus and cannot therefore properly be described as living cells. When platelets are transfused they survive only a few days and, since the concentration is greater in the splenic artery than in the splenic vein, it can reasonably be assumed that they are taken up in the reticulo-endothelial cells of the spleen. This opinion is strengthened by the finding that after splenectomy the number of platelets is increased for about two months.

The platelets may be enumerated in a counting chamber after the red cells have been removed by adding a hæmolytic agent such as urea. There are normally some 250,000 to 500,000 platelets in each cu. mm of circulating blood. The number probably fluctuates rapidly from hour to hour and it increases after exercise, hæmorrhage and surgical operation.

It is probable that platelets are group specific. Like red cells they can be agglutinated by specific antisera and, in the presence of complement, lysis can occur. When platelets disintegrate, for example during the process of coagulation, histamine and serotonin (5-hydroxytryptamine) are released. The latter substance is a potent vasoconstrictor and it has been suggested that it is responsible for the constriction of the minute vessels which is an important factor in the initial stages of hæmostasis (p. 368).

Blood platelets possess adhesive qualities by virtue of which they stick rapidly to foreign surfaces such as damaged endothelium or glass. This stickiness is increased after surgical operation, becoming maximal about the tenth day after operation, and may be a factor in the production of thrombosis in veins which sometimes occurs about that time. Platelet adhesiveness is reduced by anticoagulants.

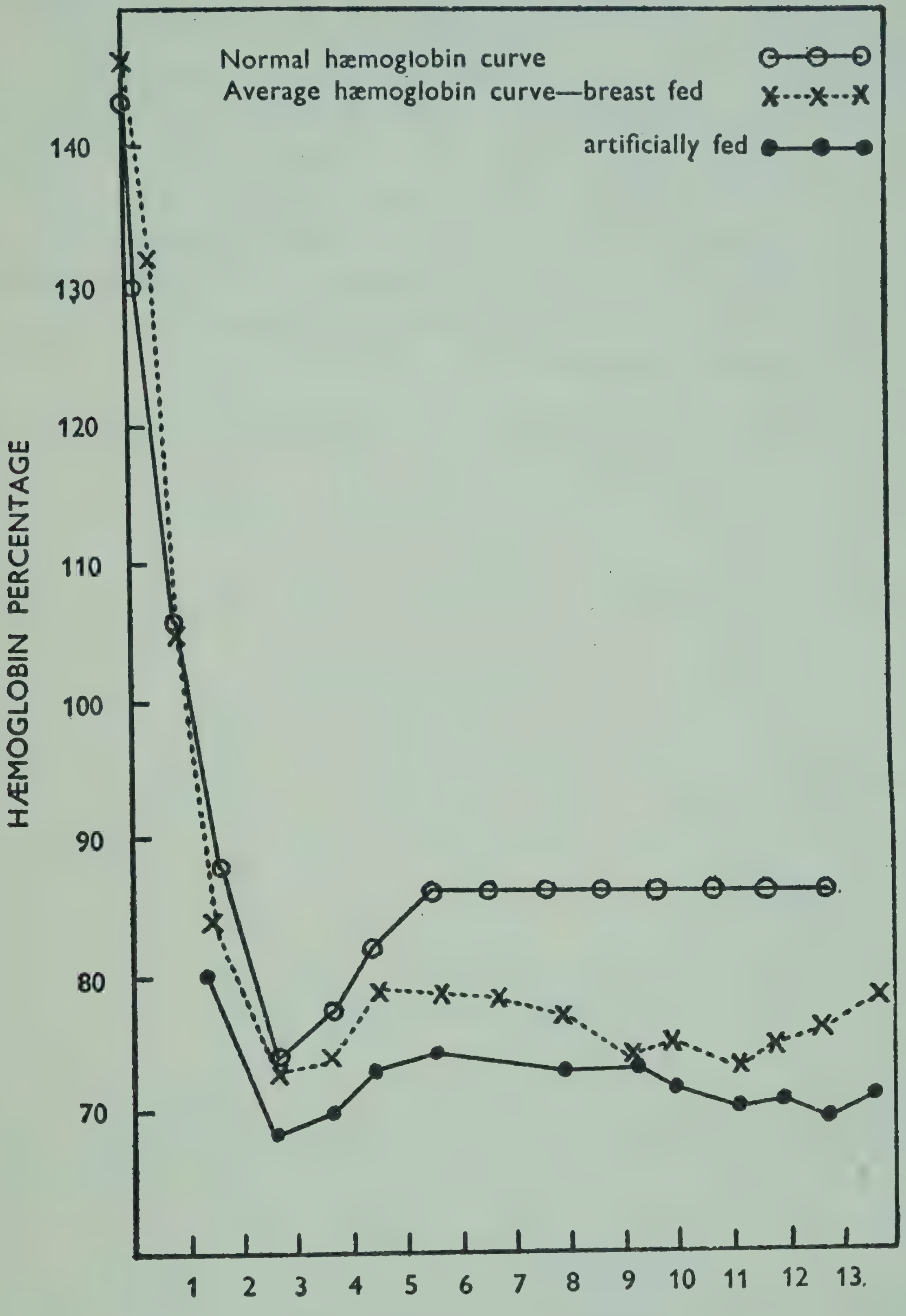


FIG. 24, 11

The 'normal' and the 'average' hæmoglobin levels during the first year of life, in terms of the Price-Jones–Haldane standard. 100 per cent. = 18·5 ml. oxygen capacity or 13·8 g. Hb per 100 ml. blood (Helen M. M. Mackay (1933), *Arch. Dis. Child.* **8,** 223).

Platelet counts below 60,000 to 80,000 per cu. mm are often associated with bleeding, although it is probable that a low platelet count is not itself the cause of hæmorrhage. The role of platelets in the processes of clot retraction and of blood clotting has already been dealt with (p. 365 to p. 367).

Reticulo-endothelial System

This is a 'system' of cells scattered throughout the body. Their phagocytic properties can easily be demonstrated if, during life, India ink or certain dyes are injected intravenously. These cells take up the particles of dye and, therefore, become stained (vital or supravital staining). They can thus be clearly seen in histological preparations, especially in the spleen, bone marrow, liver (Kupffer cells) and in the lymph nodes. There are also mobile mononuclear phagocytic cells in the tissues, usually called histiocytes. It is likely that some of these migrate into the blood stream, where they are described as monocytes. The reticulo-endothelial cells act as scavengers of tissue debris and of bacteria in infectious processes. The role of the reticulo-endothelial system in destruction of red cells and bilirubin formation is discussed on p. 414.

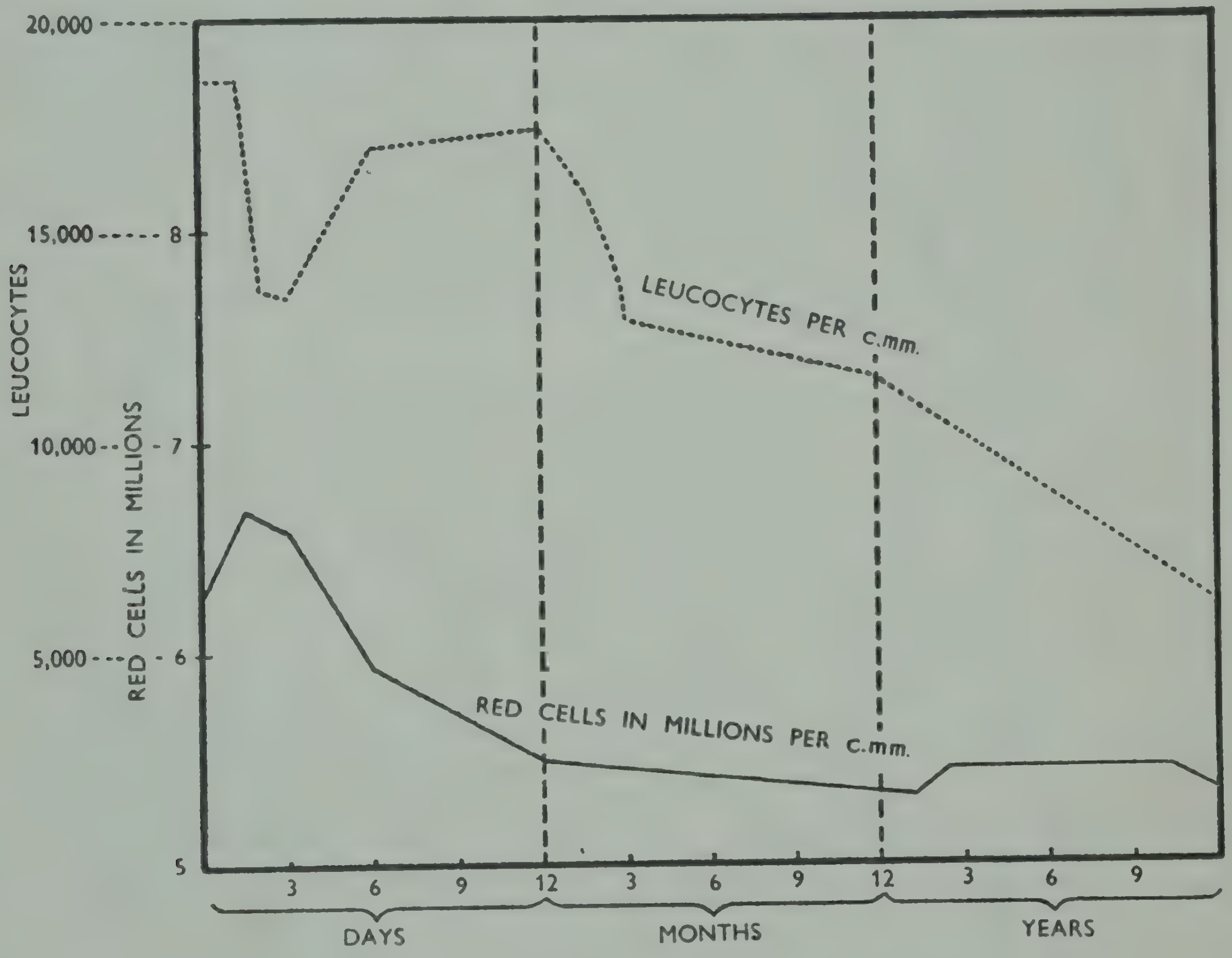


Fig. 24, 12

Red cell and leucocyte counts from birth to twelve years of age (L. E. H. Whitby & C. J. C. Britton (1953), *Disorders of the Blood*, 7th ed. p. 420. London: Churchill).

Physiological Changes in Blood in Infancy and Childhood

The hæmoglobin content of the blood at birth is about 17 g. per 100 ml., but it declines from this high value during the first two or three months of infancy to about 10 g., rising again to about 12 g. in the succeeding three months. The iron released by breakdown of the hæmoglobin of the destroyed red cells remains within the body and is re-utilized for the synthesis of hæmoglobin, but the initial fall in the red cell count is not prevented by the administration of iron even in large amounts. A proportion of the hæmoglobin of the fœtus differs both chemically and spectroscopically from the hæmoglobin of adults, the difference lying partly in the globin fraction; this so-called fœtal hæmoglobin forms as much as 80 per cent. of the total circulating hæmoglobin in the normal full-term infant but less than 1·8 per cent. in the adult. Solutions containing fœtal hæmoglobin prepared by laking blood have a lower affinity for oxygen than have similar solutions of maternal hæmoglobin. As this difference disappears after dialysis it cannot be the result of intrinsic differences in the molecule.

The red cell count at birth is about 7×10^6 per cu. mm but it declines to $5{\cdot}5 \times 10^6$ in about a fortnight. The red cells in infancy are a little larger than in adult life. A few nucleated red cells with about 5 to 10 per cent. of reticulocytes occur in the blood of the new-born but adult numbers (1 per cent. or less) are reached in a few days.

At birth there is a polymorphonuclear leucocytosis lasting only several days, followed by a lymphocytosis which persists for a year or so. The former is probably the result of deficient supply of oxygen before birth and its disappearance follows improved oxygen supply after birth. These changes are summarized in Figs. 24, 11, and 24, 12.

The platelets, unlike the other cellular elements, are present at birth in reduced numbers, normal adult values being attained within the first ten days of life.

The changes in blood volume in infancy are described on p. 376.

REFERENCES

Davis, L. J. & Brown, A. (1954). *Yhe Mebaloblastic Anæmias.* Oxford: Blackwell.

Gabler, C. J. (1956). Absorption and metabolism of iron. *Science,* **123,** 87–90.

Grant, W. C. & Root, W. S. (1952). Fundamental stimulus for erythropoiesis. *Physiol. Rev.* **32,** 449–498.

Ham, T. H. & Curtis, F. C. (1938). Sedimentation rate of erythrocytes. *Medicine, Baltimore* **17,** 447–517.

Harris, H. (1954). Role of chemotaxis in inflammation. *Physiol. Rev.* **34,** 529–562.

Mollison, P. L. (1952–3). The life span of red cells. Chap. XV of *Lect. Sci. Bas. Med.* **2,** 269–286.

Pirrie, R. (1950). Iron metabolism. *Glasg. med. J.* **31,** 397–416.

Ponder, E. (1948). *Hæmolysis and Related Phenomena.* New York: Greene & Stratton.

Tocantins, L. M. (1938). The mammalian blood platelet in health and disease. *Medicine, Baltimore* **17,** 155–260.

Whitby, L. E. H. (1954). The dynamics of hæmopoiesis. *Brit. med. J.* **2,** 1279–1284.

Whitby, L. E. H. & Britton, C. J. C. (1953). *Disorders of the Blood.* 7th ed. London: Churchill.

Yoffey, J. M. (1950). The mammalian lymphocyte. *Biol. Rev.* **25,** 314–343.

CHAPTER 25

BLOOD PIGMENTS

HÆMOGLOBIN is a conjugated protein consisting of the histone globin united to the iron-containing substance hæm. The hæmoglobins of different species vary slightly in their properties, for example some can be crystallized much more readily than others. Since hæm is common to hæmoglobins from all species these differences must be due to species variations in the globin. Hæmoglobin contains 0·34 per cent. of iron so that, if it is assumed that each molecule contains one atom of iron, the molecular weight must be 17,000. When the molecular weight is determined by the ultracentrifuge, however, a value of 68,000 is found and it is therefore likely that the hæmoglobin molecule contains four atoms of iron, being composed of four units, each of molecular weight 17,000 and each containing one atom of iron.

The most important property of hæmoglobin is its power of combining with oxygen to form an easily dissociated compound known as oxyhæmoglobin. When hæmoglobin is exposed to oxygen or air, it takes up two atoms of oxygen for each atom of iron. If we represent the hæmoglobin molecule as Hb to indicate the unit of molecular weight 17,000 then the combination with oxygen can be represented thus:

$$\mathrm{Hb} + \mathrm{O_2} \rightleftharpoons \mathrm{HbO_2}$$

If we represent the complete hæmoglobin molecule of molecular weight 68,000 and containing four atoms of iron as HB then the combination with oxygen could be represented by the equation:

$$\mathrm{HB} + 4\mathrm{O_2} \rightleftharpoons \mathrm{HB[O_2]_4}$$

This combination with oxygen is not an oxidation, it is simply a loose union between hæmoglobin and oxygen. No change in valency occurs in the formation of oxyhæmoglobin; the iron is in the ferrous state in both hæmoglobin and oxyhæmoglobin. Oxyhæmoglobin gives up its oxygen very readily, when, for example, blood is exposed to conditions of low oxygen tension.

$$\mathrm{HbO_2} \longrightarrow \mathrm{Hb} + \mathrm{O_2}$$

The ease with which hæmoglobin unites with oxygen and with which the oxyhæmoglobin gives up its oxygen again is, as we shall see later, of the utmost importance in the carriage of oxygen by the blood (p. 573).

Hæmoglobin can combine with gases other than oxygen to form addition compounds. For example, it combines reversibly with carbon monoxide forming carboxyhæmoglobin (carbonyl hæmoglobin, carbon monoxide hæmoglobin or COHb).

$$\mathrm{Hb} + \mathrm{CO} \rightleftharpoons \mathrm{COHb}$$

Not only is this compound more stable than oxyhæmoglobin but the affinity of hæmoglobin for carbon monoxide is some 250 times greater than the affinity of hæmoglobin for oxygen. For this reason even small concentrations of carbon monoxide in the inspired air are dangerous, since the carboxyhæmoglobin deprives the blood of its ability to transport oxygen and the tissues are rapidly asphyxiated. The breathing of so small a concentration as 0·06 vol. per cent. carbon monoxide in air produces symptoms within an hour and unconsciousness in two hours. Even in non-fatal cases of carbon monoxide poisoning, permanent damage to the central nervous system, the cells of which are particularly sensitive to lack of oxygen, may be produced. Since the formation of carboxyhæmoglobin is a reversible reaction it is possible to displace carbon monoxide by exposure of carboxyhæmoglobin to oxygen for a considerable time. This fact is made use of in the treatment of persons suffering from carbon monoxide poisoning due to the inhalation of coal gases or of exhaust gases from internal combustion engines. Since the rate at which oxygen displaces carbon monoxide is proportional to the pressure of oxygen (law of mass action) it is advisable to administer pure oxygen to such persons (see p. 586). Carboxyhæmoglobin has a characteristic cherry-pink colour and patients suffering from severe carbon monoxide poisoning may have a strikingly pink complexion even when their oxygen-carrying capacity is impaired to the point of asphyxia.

Up to 5 per cent. of the hæmoglobin in the blood of heavy tobacco smokers may be in the form of carboxyhæmoglobin.

METHODS OF ESTIMATION OF HÆMOGLOBIN

Accurate measurements of the concentration of hæmoglobin in blood may be made in two ways, namely by measurement of the iron content of a sample of blood, or by determining its oxygen content after it has been fully saturated with oxygen. It might be thought that the latter method, since it measures the physiologically available hæmoglobin, would be the better. Many blood samples, however, contain compounds such as methæmoglobin and carboxyhæmoglobin which do not combine with oxygen. For these reasons the iron method is now preferred. The methods of measuring the oxygen capacity of blood are dealt with on p. 572.

Estimation of the concentration of hæmoglobin in the blood is so important in clinical medicine that it is desirable to have some relatively simple and rapid method for its determination, even if this should mean some sacrifice of accuracy. As might be expected, the principle underlying all such methods is colour comparison, the colour of the patient's blood, usually after some chemical treatment, being compared with a standard colour. Because of the difficulty of obtaining a satisfactorily pure sample of human hæmoglobin from which a standard could be prepared, various compromises have been effected.

One of the most widely used was introduced by Haldane. He took samples of blood from a number of healthy people, mixed them, converted the hæmoglobin into carboxyhæmoglobin by means of coal-gas in order to facilitate colour comparison, made a 1 in 100 dilution of the blood with water and arbitrarily defined the resulting colour as representing 100 per cent. hæmoglobin. The colour values of this standard have recently been defined by the National Physical Laboratory and iron analyses have shown that 100 per cent. on this accurately characterized standard represents

14·8 g. hæmoglobin per 100 ml. blood, corresponding to an oxygen capacity of 19·8 ml. per 100 ml. blood or 49 mg. hæmoglobin iron per 100 ml. blood.

There are several practical difficulties associated with the use of the Haldane standard. One is that the colour of the standard tends to fade. Another is that the colour comparison is a difficult one to make and a third is the use of coal gas with which it is necessary to treat every sample of blood before the comparison can be made. To overcome these objections subsequent workers have converted the hæmoglobin into acid or alkaline hæmatin and have prepared coloured glass standards, the intensity of colour of which represents 100 per cent. Many of these differ from the Haldane standard in value, one of the commonest (Sahli) being equivalent to 17·3 g. hæmoglobin per 100 ml. blood. More recently various rapid and accurate methods have been devised for clinical use based upon pure crystalline hæmin as standard from which a solution is prepared whose colour is the equivalent of 14·8 g. per 100 ml. blood. Both the standard solution and the blood whose hæmoglobin concentration it is desired to determine are treated to convert the hæmoglobin into either alkaline hæmatin or cyan-hæmatin. These methods measure all the hæmoglobin, including oxyhæmoglobin, hæmoglobin and methæmoglobin.

If a photoelectric colorimeter is available then practically any derivative of hæmoglobin including oxyhæmoglobin may be used to estimate very rapidly the concentration with a much greater accuracy than is needed for clinical work. The error may be less than 1 per cent. The diurnal variations in hæmoglobin are probably greater than this. Photoelectric colorimeters, better called absorptiometers, measure very quickly and accurately the ratio of the light incident on a solution of hæmoglobin to the light emerging from the solution in a narrow band of the spectrum. The logarithm of this ratio, called the optical density of the solution, is proportional to the hæmoglobin percentage (Beer's law).

Probably the simplest method is to measure, using a green filter, the density of a dilution of the unknown blood sample in a very weak alkaline solution. The hæmoglobin then takes up oxygen very quickly from the water and is converted completely to oxyhæmoglobin. The instrument is calibrated by measuring the optical density of blood samples of known iron content or oxygen capacity. Hæmoglobinometers designed by a committee of the Medical Research Council and using this principle are now available. The observer looks through a green filter at the oxyhæmoglobin solution and at a calibrated grey wedge which is adjusted until the two colour intensities match.

There is considerable variation between the values obtained for the concentration of hæmoglobin in the blood of normal healthy adults. In Britain, the normal range for adult males may be taken as 14 to 17 g. per 100 ml. of blood with a mean at 15·6 g. In the female the range is from 12 to 15·5 g. per 100 ml. with a mean at 13·7 g.

In the investigation of patients with anæmia the *colour index* is often determined. This index expresses the mean hæmoglobin content of a single cell relative to the content of a normal cell thus:—

$$\text{Colour index} = \frac{\text{Hæmoglobin concentration as percentage of normal}}{\text{Number of red cells as percentage of normal}}$$

In calculating the colour index 14·8 g. of hæmoglobin per 100 ml. is regarded as 100 per cent. hæmoglobin and 5,000,000 red blood cells per cmm. as 100 per cent. in the denominator. The normal colour index is clearly 1, although figures between 0·85 and 1·15 cannot be regarded as abnormal.

It will be evident that the size of the red cells is sometimes an important factor influencing the value of the colour index because the amount of hæmoglobin in a single red cell relative to the normal may be determined by the size of the cell. The colour index is thus unsatisfactory as a measure of hæmoglobin load, depending as it does, partly on cell size and partly on the saturation of the stroma of the cell with hæmoglobin.

The mean corpuscular hæmoglobin concentration, that is the mean concentration of hæmoglobin per 100 ml. of cells, is an index that is independent of the size of the red cells and is therefore a true expression of their hæmoglobin load.

$$\text{M.C.H.C.} = \frac{\text{Hæmoglobin concentration (g. per 100 ml.)}}{\text{Vol. of red cells in 100 ml. of blood (hæmatocrit)}} \times 100$$

Normal values lie between 32 and 38 g. per 100 ml. Figures below this range indicate that the stroma of the red blood cells is not carrying its full quota of hæmoglobin and strongly suggest that the patient will benefit from the administration of iron. Since the normal hæmoglobin load, as expressed by an M.C.H.C. between 32 and 38 g. per 100 ml., represents the maximum that can be carried, values above 38 g. per 100 ml. cells are never found.

Tests for Blood

The presence of blood in biological material or in suspected blood stains may be detected by a number of methods.

1. **Colour Tests.**—These tests depend on the fact that hæmoglobin and its derivatives act as peroxidases which are stable to heat. Thus if a solution suspected to contain blood is boiled and cooled, and is then treated with a small amount of hydrogen peroxide and either benzidine or guaiaconic acid, a blue oxidation product is obtained if hæmoglobin is present. This is a very delicate test for hæmoglobin but it is given by a number of other organic compounds containing iron though not by inorganic salts.

2. **Spectroscopic Methods.**—Hæmoglobin and its derivatives have very characteristic absorption spectra (Fig. 25, 2), which allow of their detection and recognition in considerable dilutions. For example, a suspected blood stain could be treated so as to produce a particular hæmochromogen and the spectrum obtained would at once indicate whether the pigment had been derived from blood.

3. **The Production of Crystals of Hæmin or of Pyridine Hæmochromogen.**—If a small sample of material suspected to contain dried blood is heated on a microscope slide with a small amount of glacial acetic acid and sodium chloride, characteristic crystals of hæmin (see p. 410) are formed and may be detected microscopically. Alternatively a fragment of the blood stain may be heated on a microscope slide with a small amount of pyridine, sodium hydroxide and a reducing agent such as glucose or sodium hydrosulphite, when salmon-pink needles of pyridine hæmochromogen are formed, the characteristic form of which may be recognized under the microscope (Takayama's test).

The Absorption Spectra of Blood Pigments (Fig. 25, 2)

Hæmoglobin and its derivatives may readily be distinguished from one another by means of their absorption spectra. If a dilute solution of hæmoglobin (1: 50 blood) containing a small amount of reducing agent is placed before the slit of a spectroscope a single broad absorption band is seen between the red and the green, with its centre at 565 mμ. Oxyhæmoglobin gives two narrower bands also in the red-green, an α band at 578 mμ and a β band at 540 mμ. Carboxyhæmoglobin gives two very similar bands at 572 mμ and 535 mμ. Since the displacement of the two bands of carboxyhæmoglobin towards the blue end of the spectrum is so small it is difficult to distinguish between oxyhæmoglobin and carboxyhæmoglobin, although it is obviously important to be able to do so.

To overcome this difficulty Hartridge devised the reversion spectroscope (Fig. 25, 2) which shows two spectra one above the other, the lower one running in the opposite direction to the upper one, so that the red end of the upper spectrum is on the left and the red end of the lower spectrum is on the right. By means of a screw the upper spectrum can be moved laterally over the lower. If oxyhæmoglobin is examined in the instrument, the relative positions of the two spectra can be adjusted so that the α band of the upper spectrum is exactly over the α band of the lower spectrum. The oxyhæmoglobin is then replaced with a solution suspected to contain carboxyhæmoglobin. If carboxyhæmoglobin is present the band in the upper spectrum will be shifted slightly to the right and in the lower spectrum slightly to the left so that the bands no longer coincide.

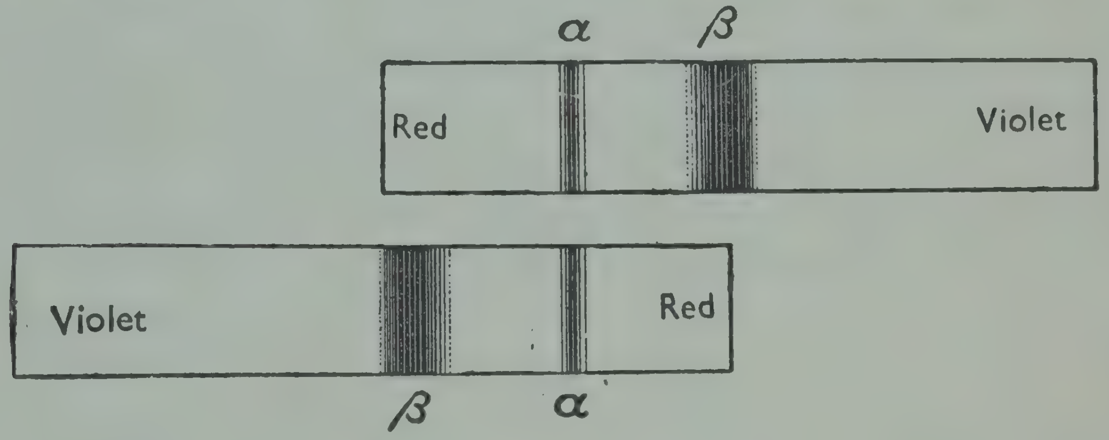


Fig. 25, 1

Diagrammatic representation of the two spectra of oxyhæmoglobin as seen in the reversion spectroscope devised by Hartridge.

Pigments related to Hæmoglobin

Methæmoglobin.—When hæmoglobin is treated with mild oxidizing agents such as potassium chlorate or potassium ferricyanide it becomes oxidized to a stable oxidation product known as methæmoglobin. In this process one atom of oxygen is taken up for each atom of iron.

$$\mathrm{HbFe^{++}O_2 + Fe(CN)_6^{\equiv} \longrightarrow HbFe^{+++} + Fe(CN_6)^{\equiv} + O_2}$$

or more simply

$$\mathrm{Hb + \tfrac{1}{2}O_2 \longrightarrow HbO}$$

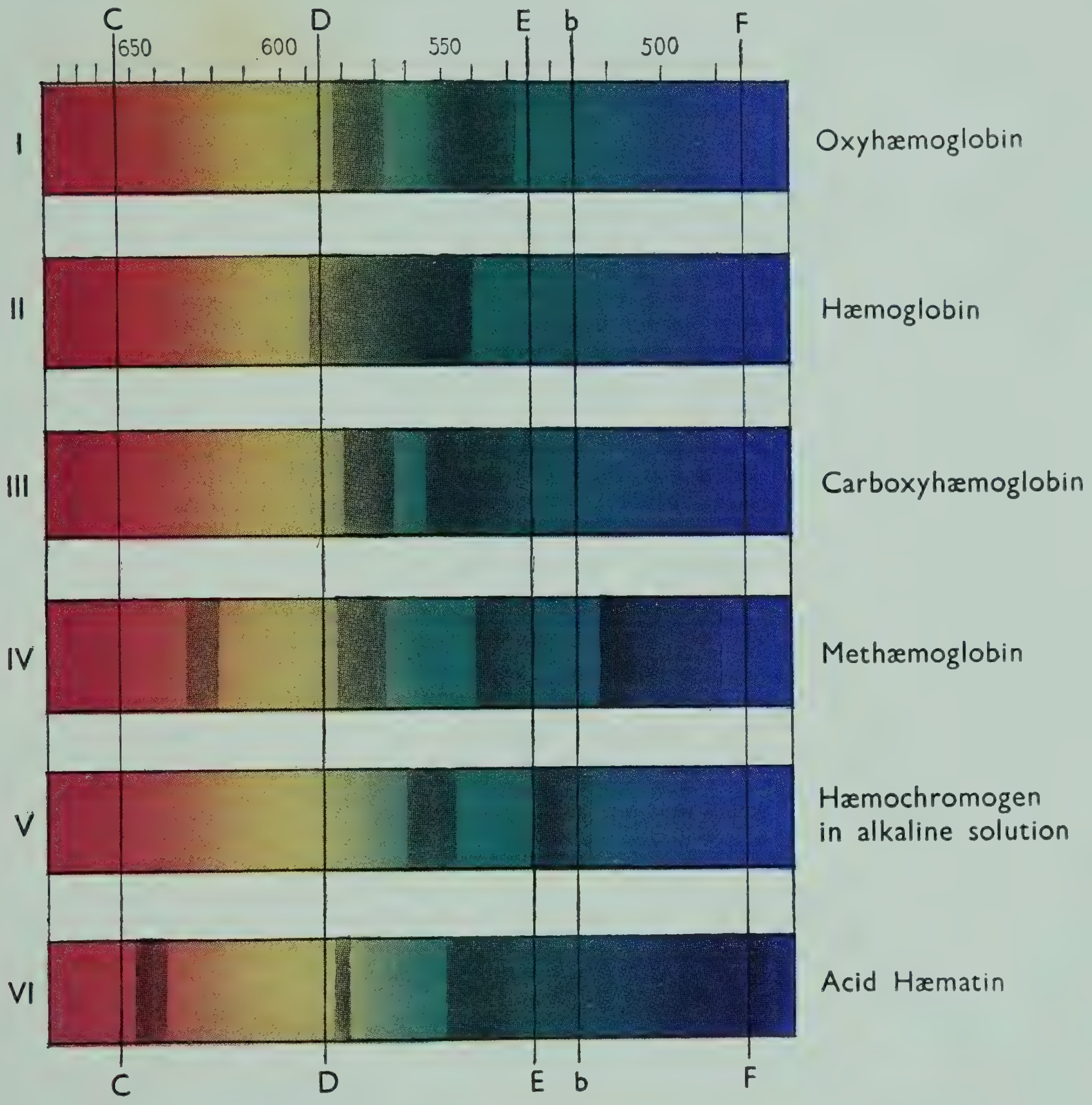


FIG. 25, 2

Absorption spectra of some hæmoglobin derivatives. The scale at the top shows the wavelength in mμ. The letters denote the position of the Fraunhofer lines of the solar spectrum.

Since the loosely combined molecular oxygen in HbO_2 is driven off, this reaction with ferricyanide is used to determine the oxygen capacity of blood (p. 572). Whereas oxyhæmoglobin is merely a loose association of hæmoglobin with molecular oxygen, methæmoglobin is a true oxidation product. In both hæmoglobin and oxyhæmoglobin iron is present in the ferrous state, while in methæmoglobin iron is in the ferric state. Methæmoglobin is occasionally found in the blood after the administration of drugs as phenacetin and some of the sulphonamides, and rarely as a familial anomaly. The condition is known as methæmoglobinæmia. It has been estimated that from 2 to 12 per cent. of the hæmoglobin in the blood of healthy persons may be present in an 'inactive' form, probably as methæmoglobin. Methæmoglobin may be reduced in vitro to hæmoglobin by strong reducing agents such as sodium hydrosulphite ($Na_2S_2O_4$) (Fig. 25, 3). Since methæmoglobin can be reduced to hæmoglobin in vivo by ascorbic acid this substance is useful in the treatment of methæmoglobinæmia.

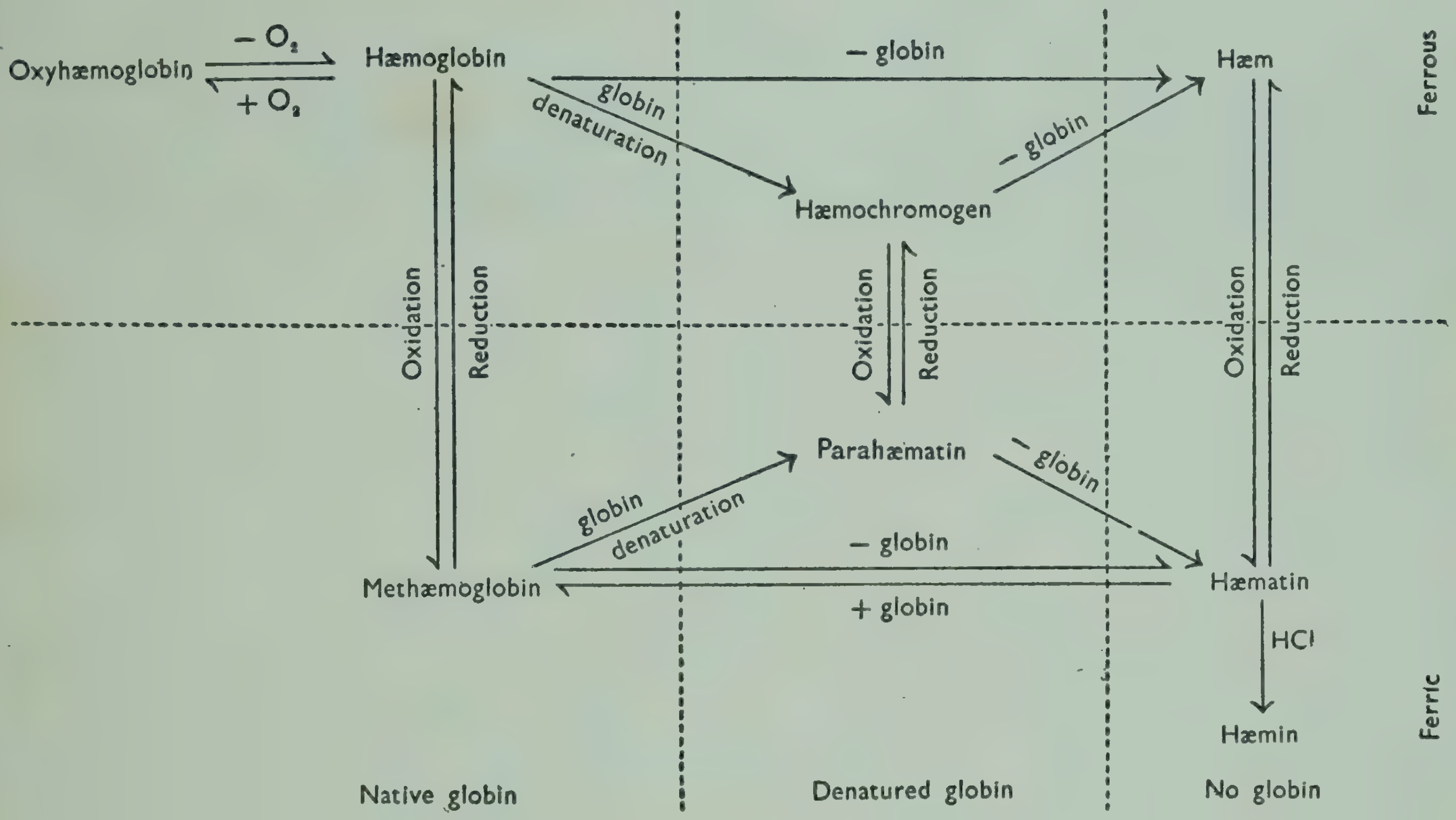

Oxyhæmoglobin
− O₂
+ O₂
Hæmoglobin
− globin
globin denaturation
− globin
Hæm
Ferrous
Hæmochromogen
Oxidation
Reduction
Oxidation
Reduction
Oxidation
Reduction
Parahæmatin
globin denaturation
− globin
− globin
Methæmoglobin
+ globin
Hæmatin
HCl
Hæmin
Ferric
Native globin
Denatured globin
No globin

FIG. 25, 3

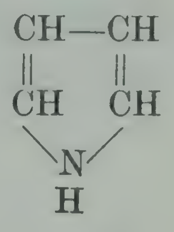
Relationships of the blood pigments.

Another abnormal pigment occasionally found in the blood is *methæmalbumin* in which the hæm is combined with plasma albumin. Its presence in the plasma may be an important indication of a hæmolytic process.

Sulphæmoglobin may be found in the blood after administration of certain drugs. It is a stable compound which, unlike methæmoglobin, cannot be reduced to hæmoglobin. Sulphæmoglobin remains as such in the red cells during their life span.

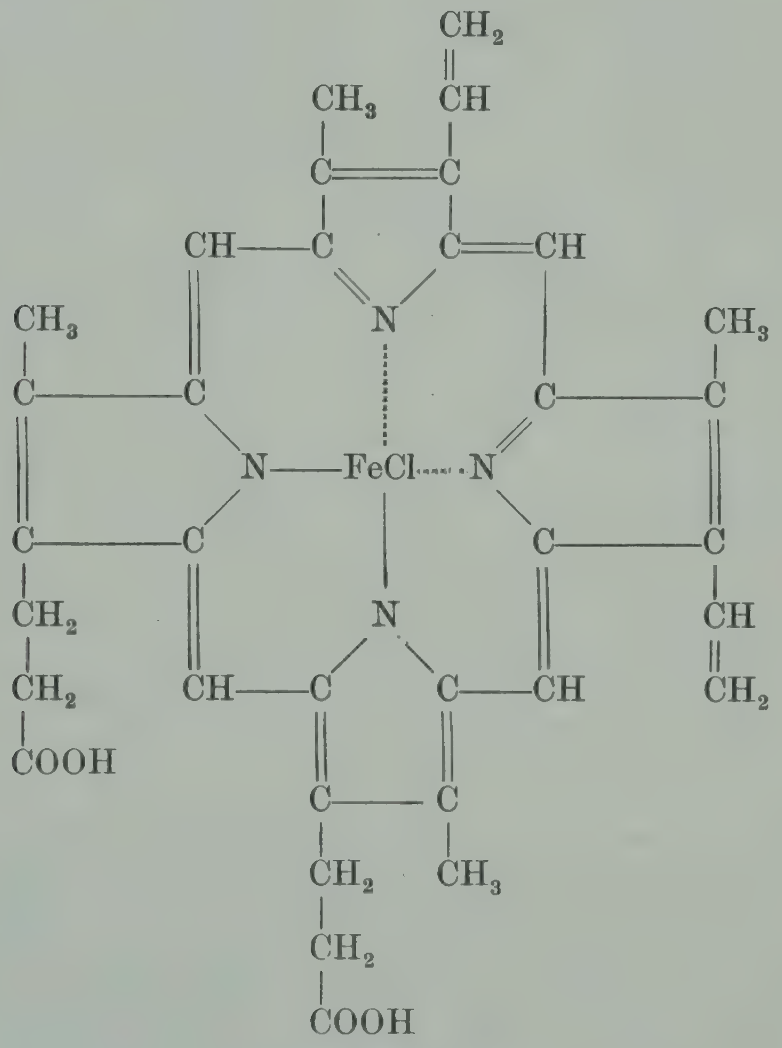
Hæm and Hæmatin.—Hæmoglobin, as already mentioned, consists of the pigment hæm bound to the protein globin. Hæm, like hæmoglobin, contains iron in the ferrous state and it is easily oxidized to the corresponding compound *hæmatin* which contains ferric iron. Hæmatin reacts readily with hydrochloric acid to form a hydrochloride which is known as *hæmin*. Hæmin crystallizes easily and the formation of hæmin crystals, which can be easily recognized under the microscope, is the basis of a well-known test for blood (p. 407). Hæmatin can be obtained not only from hæmoglobin but also from other biological materials such as myoglobin (muscle hæmoglobin), the enzymes catalase and peroxidase, and from cytochrome.

The hæmin molecule is composed of four pyrrol rings,



pyrrol

united by four CH groups to form a flat closed ring system. The iron atom occupies the centre of this large ring and is united to the nitrogen atoms of the pyrrol rings (Fig. 25, 4).



hæmin

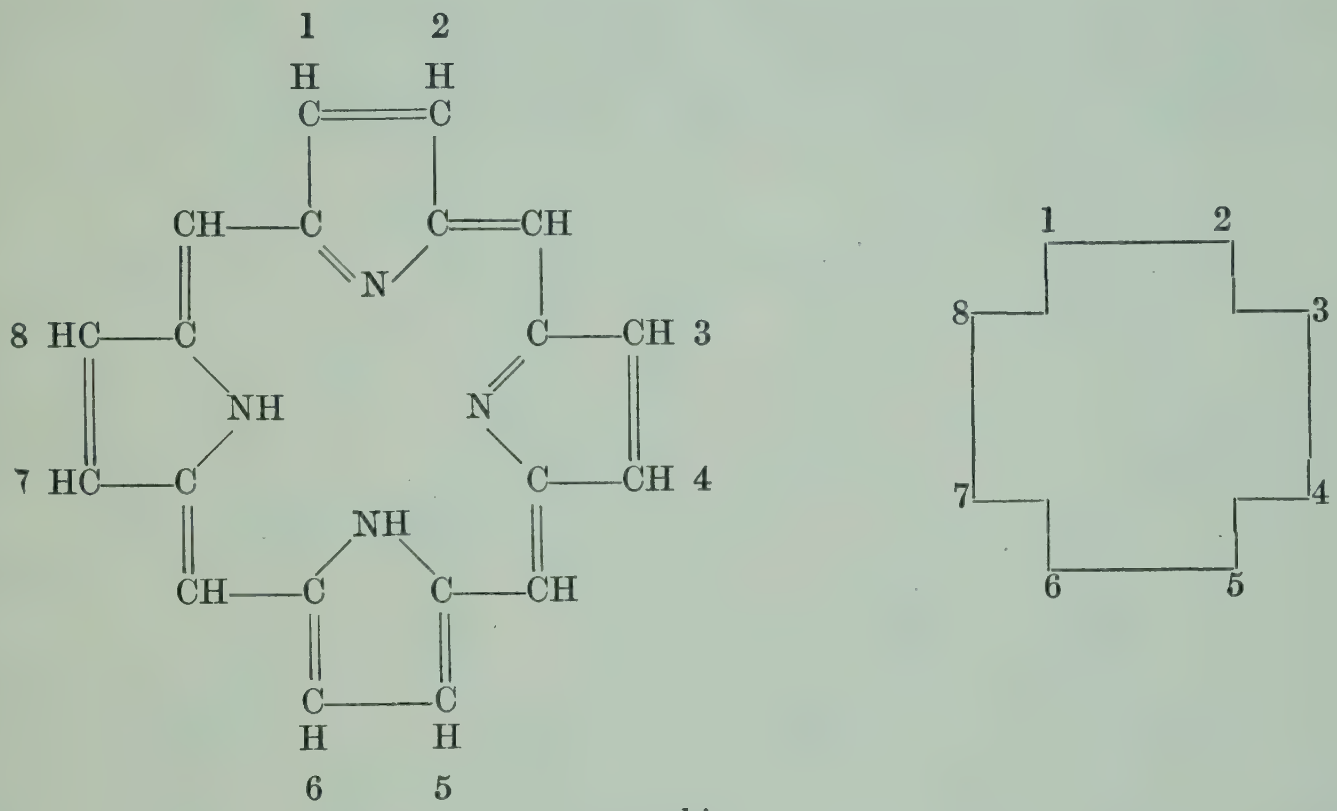
FIG. 25, 4

The carbon atoms of the pyrrol rings carry substituent groups which are either methyl (CH_3–), vinyl (CH_2=CH–), or propionic acid (HOOC.CH_2.CH_2–) groupings.

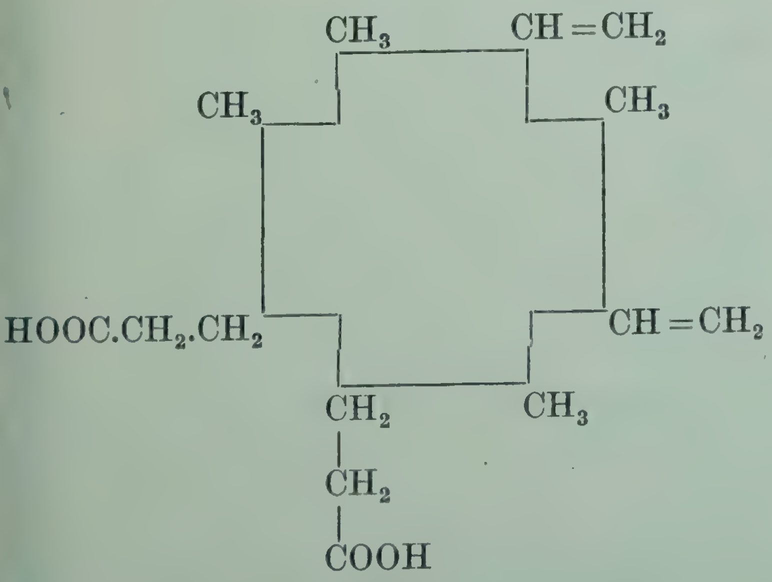
Hæmochromogens.—When hæmoglobin is treated with alkali the protein is denatured but it remains attached to the hæm, and the compound formed, consisting of hæm united to denatured globin, is known as *hæmochromogen* or, more correctly, *globin hæmochromogen*. In the preparation of hæmochromogen a small amount of reducing agent is usually added to ensure that the iron remains in the ferrous state. The compound corresponding to hæmochromogen in which the iron is in the ferric state is known as *parahœmatin*, and it may be formed either by oxidizing hæmochromogen or by denaturing the globin of methæmoglobin (see Fig. 25, 3).

The term hæmochromogen is a general one, and is applied to any substance formed by the combination of hæm with a nitrogenous substance, which may be denatured globin as in the case of globin hæmochromogen, or some such compound as pyridine, ammonia, or glycine. Pyridine hæmochromogen is readily formed as characteristic crystals and is frequently prepared in the identification of blood stains (p. 407). Cytochrome is another example of a naturally occurring hæmochromogen. It has been discussed on p. 83.

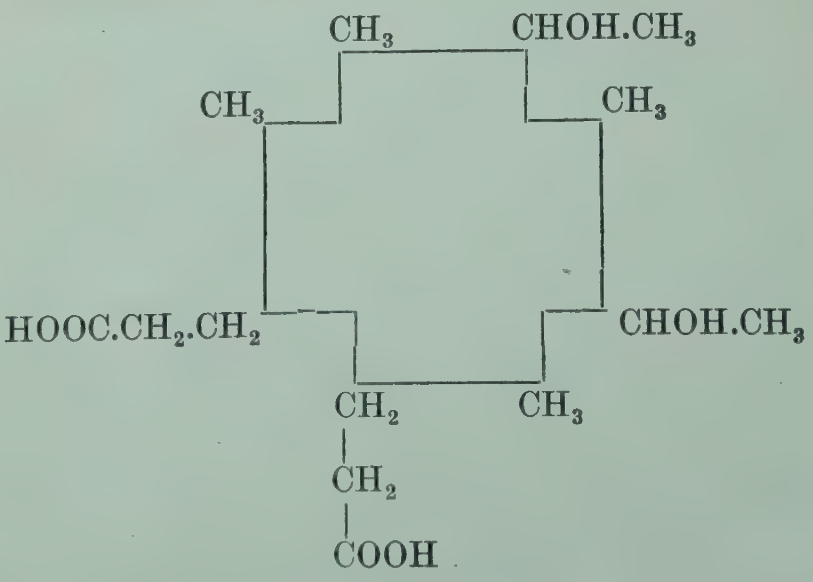
Porphyrins.—When iron is removed from hæm or hæmatin we are left with the ring system composed of four pyrrol nuclei joined together by the CH groupings. A compound of this sort is known as *porphyrin* and all porphyrins are derivatives of the parent substance *porphin*, the full formula of which is shown on the left. The many porphyrins are characterized by different substituent groupings in positions 1 to 8 in the porphin ring



porphin



protoporphyrin



hæmatoporphyrin

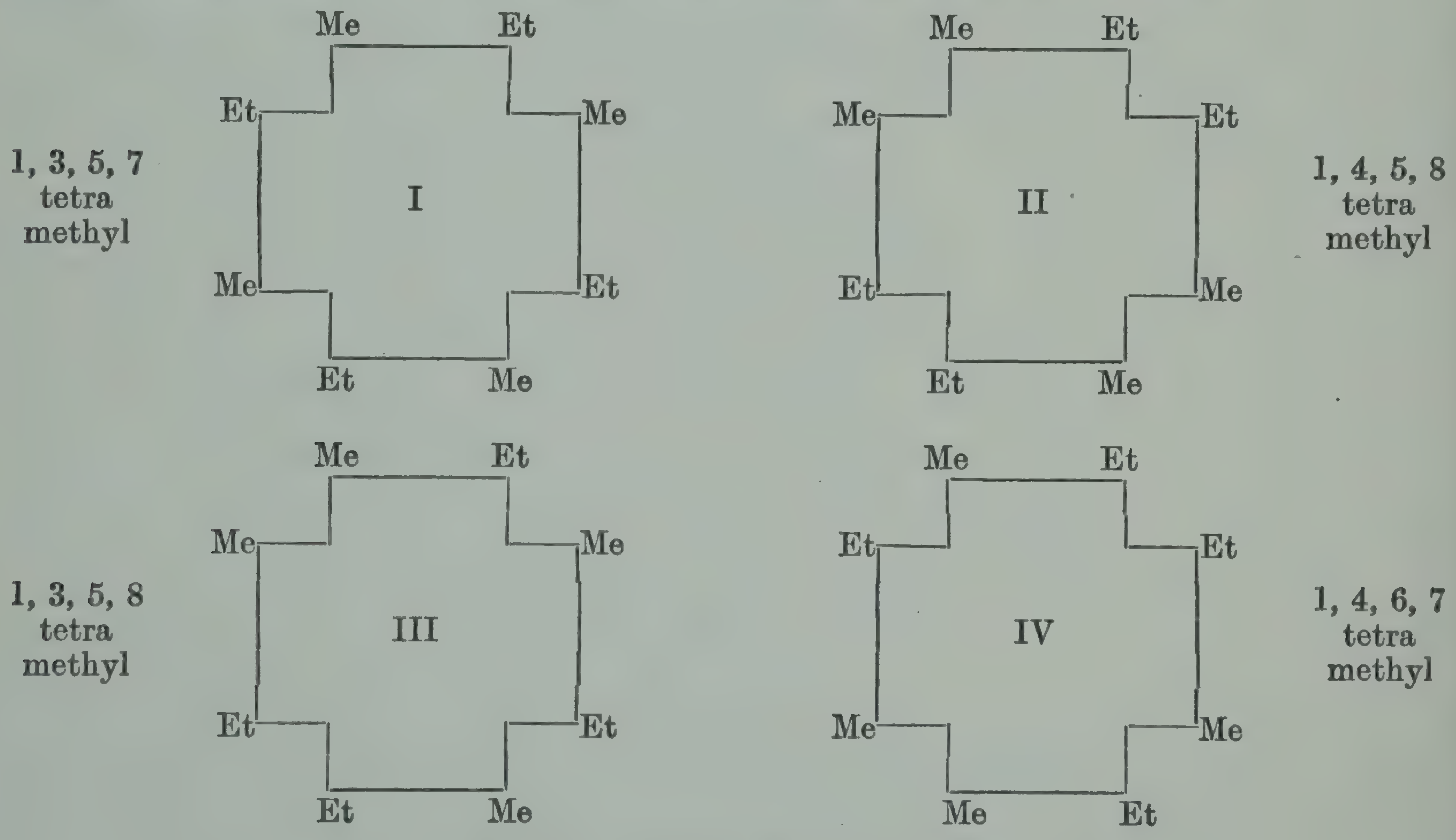
system. To save time and space the porphin ring is frequently shown in diagrammatic form as on the right, where each corner of the figure represents the point of attachment of the substituent groups in the pyrrol rings. The porphyrin which is obtained from hæmatin or hæm is known as *protoporphyrin* and has methyl groups attached in positions 1, 3, 5 and 8, vinyl groups at positions 2 and 4, and propionic acid groupings at positions 6 and 7.

Protoporphyrin can be prepared from hæmoglobin or hæmatin by treatment with strong acid. Hæmatoporphyrin is a derivative of protoporphyrin in which the vinyl groups have taken up water thus:

$$—CH=CH_2 + H_2O \longrightarrow —CHOH—CH_3$$

Hæmatoporphyrin does not occur naturally but the name has sometimes been erroneously used in medical literature to indicate the porphyrin found in urine or fæces. The pigments excreted are in fact as described below.

If hæmoglobin or hæmatoporphyrin is heated with soda-lime a porphyrin known as ætioporphyrin is formed in which the substituents are methyl or ethyl groupings only. There are four possible ætioporphyrins.



The four possible ætioporphyrins

In ætioporphyrin of type I the methyl groups are in positions 1, 3, 5 and 7; in type II the methyl groups are in positions 1, 4, 5 and 8; in type III they are in positions 1, 3, 5 and 8; and in type IV they are in positions 1, 4, 6 and 7. The ætioporphyrin derived from protoporphyrin belongs to series III, and it is now known that the naturally occurring porphyrins belong either to series I or to series III only. The urine normally contains 10 to 100 μg. of porphyrin per day and 150 to 300 μg. are excreted per day in the fæces. Since these porphyrins belong to series I as well as to series III they cannot be formed by the breakdown of hæmoglobin; they are probably derived from ingested plant and animal pigments or are by-products of hæmoglobin formation.

Increased urinary elimination of type III porphyrins and of a colourless precursor, *porphobilinogen*, may occur in poisoning with heavy metals and certain drugs and in chronic disease of the liver. It is also a feature of the grave condition known as acute porphyria which is a familial disorder inherited as a Mendelian dominant (Chapter 54). Large amounts of type I porphyrins are excreted by subjects with *congenital porphyria*, which is an inborn error of porphyrin metabolism inherited as a Mendelian recessive (Chapter 54).

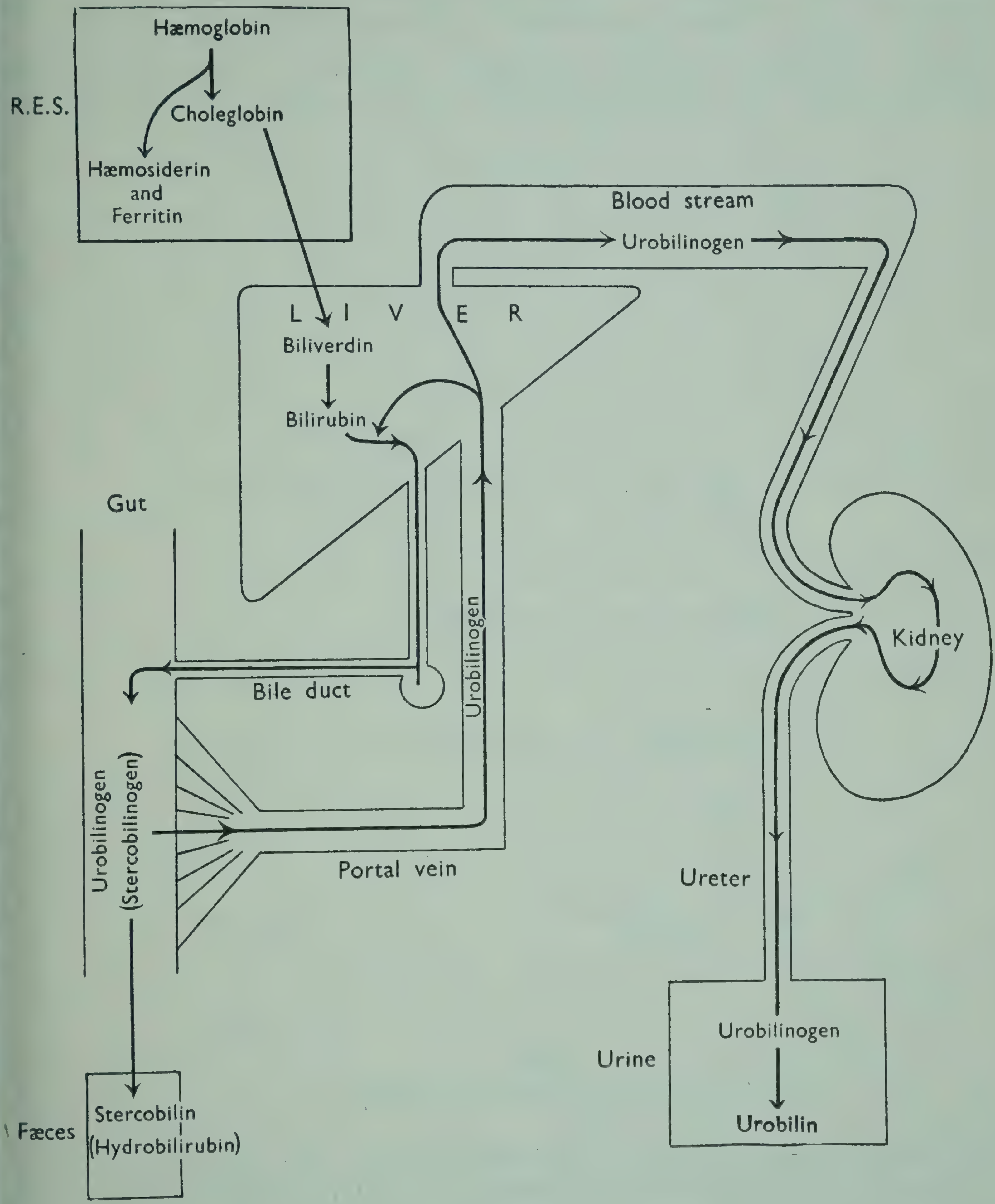

Hæmoglobin
R.E.S.
Choleglobin
Hæmosiderin and Ferritin
Blood stream
Urobilinogen
L I V E R
Biliverdin
Bilirubin
Gut
Urobilinogen
Kidney
Bile duct
Urobilinogen (Stercobilinogen)
Portal vein
Ureter
Urobilinogen
Urine
Urobilin
Fæces
Stercobilin (Hydrobilirubin)

Fig. 25, 5

Schematic representation of the fate of hæmoglobin breakdown products.

Porphyrins have the power of combining with metals. Thus protoporphyrins can combine with iron to give hæm or hæmatin, and blood pigments are therefore iron porphyrin derivatives. Chlorophyll is a magnesium porphyrin in which complex substituent groups are attached in positions 1 to 8, but on distillation with soda-lime chlorophyll, like hæmatin, yields ætioporphyrin III.

THE BREAKDOWN OF HÆMOGLOBIN

The red blood cells have a limited existence in the body (p. 390). Corpuscles are normally removed from the circulation by the reticulo-endothelial system (Figs. 25, 5, and 25, 6) to the extent that some 7 to 8 g. of hæmoglobin are taken out of circulation per day. The first stage in the degradation of the hæmoglobin molecule consists in the opening of the porphyrin ring system. The resulting compound, known as *choleglobin*, still contains iron and protein but the four pyrrol nuclei now form a chain instead of a ring. In the next stage the molecule loses both protein and iron. The latter is retained in the reticulo-endothelial cells in the pigments hæmosiderin and ferritin so that in diseases associated with excessive blood destruction the amount of hæmosiderin in the spleen, liver and bone marrow is often much increased. The iron derived from the original

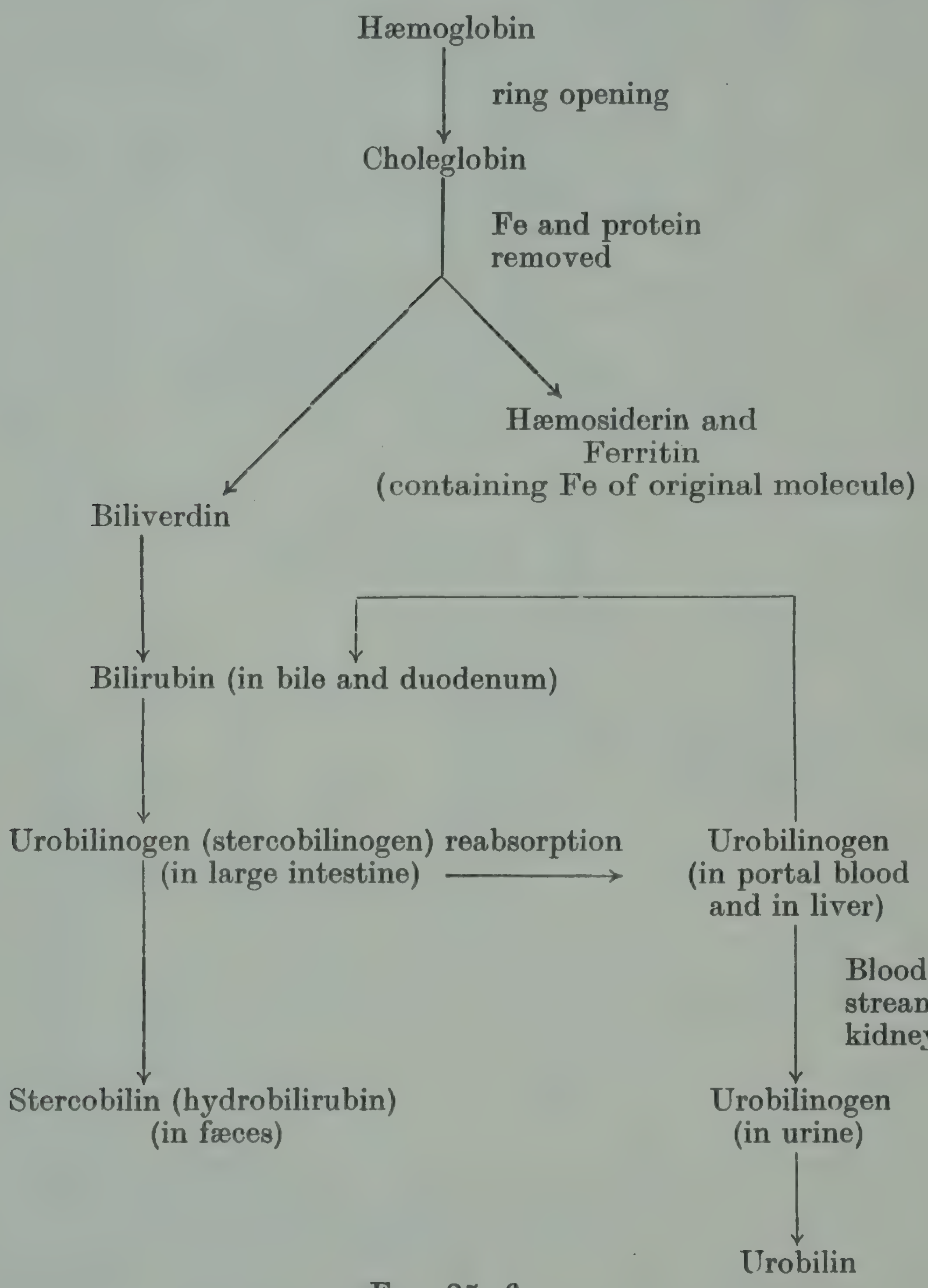
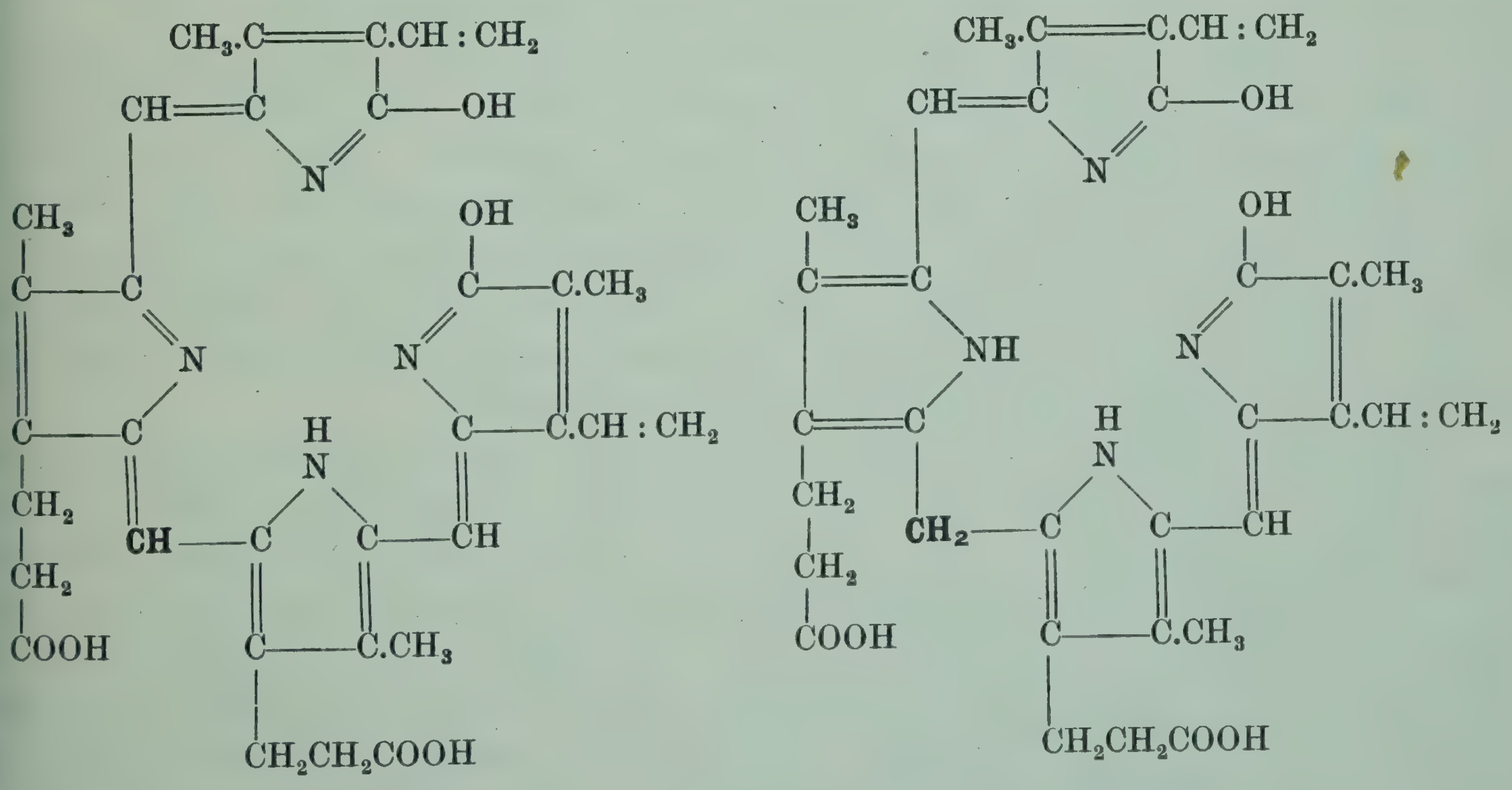

Hæmoglobin
ring opening
Choleglobin
Fe and protein removed
Hæmosiderin and Ferritin (containing Fe of original molecule)
Biliverdin
Bilirubin (in bile and duodenum)
Urobilinogen (stercobilinogen) (in large intestine)
reabsorption
Urobilinogen (in portal blood and in liver)
Blood stream, kidney
Stercobilin (hydrobilirubin) (in fæces)
Urobilinogen (in urine)
Urobilin

FIG. 25, 6

The Breakdown of Hæmoglobin

hæmoglobin molecule must be used again in the synthesis of new hæmoglobin (p. 394) since a daily turnover of 8 g. Hb corresponds to 27 mg. Fe whereas the daily loss from the body may be estimated to be 1 mg. (p. 98). After iron and protein have been removed in this way from the hæmoglobin molecule, the pigment remaining, known as *biliverdin*, is carried by the blood to the liver. Biliverdin consists of four pyrrol nuclei joined together in the form of a chain (Fig. 25, 7). Its structure is essentially that of protoporphyrin with the large ring system broken open. Some of the biliverdin, green in colour, is then reduced to the reddish-yellow *bilirubin* (Fig. 25, 7). These two compounds are the chief pigments of the bile which is secreted by the liver cells and passes to the gall-bladder and then into the duodenum. In the upper part of the intestine both pigments are reduced to *urobilinogen* (*stercobilinogen*) which in turn is converted into *stercobilin*, one of the characteristic brown pigments of the fæces normally eliminated in amounts of between 40 and 280 mg. per day. During its passage down the alimentary tract urobilinogen is absorbed into the blood stream and passes by way of the portal vein to the liver where some of it is re-excreted in the bile. A small amount of urobilinogen however passes through the liver into the general circulation and so to the kidney which excretes it in the urine. Freshly passed urine, therefore, normally contains a small amount of urobilinogen (0·5 to 2·0 mg. per day) which on standing is oxidized by atmospheric oxygen to *urobilin* (Fig. 25, 6).

With this description of the formation of these pigments certain disturbances produced by disease can be readily understood. If the common bile duct is obstructed by a stone or a tumour then bile pigment cannot pass into the intestine and the fæces are pale in colour (clay-coloured). The bile pigments are dammed back by the obstruction and are reabsorbed into the blood. The concentration of bilirubin in the blood and in the tissues rises so high that this pigment becomes easily visible, first in the sclera of the eye and later in the skin. The condition is known as *jaundice* and this kind of jaundice is said to be *obstructive*. The high level of bile



biliverdin

bilirubin

FIG. 25, 7

pigments in the blood results in their excretion by the kidney, and the urine becomes dark greenish-yellow in colour.

On the other hand there are several conditions associated with an excessive destruction of red blood cells, and consequently of hæmoglobin, in which excess bile pigment is produced and excreted into the intestine. The amount of stercobilin in the fæces is therefore increased and the fæces become dark brown, sometimes almost black, in colour. More urobilinogen than normal is reabsorbed from the gut and the concentration of urobilinogen in the urine is increased from a barely detectable level to an amount which gives strong positive tests. Thus, in hæmolytic anæmia, in which there is excessive breakdown of hæmoglobin, the urine gives strong positive tests for urobilinogen when fresh, and for urobilin after standing. Should the production of bile pigment exceed the capacity of the liver to excrete it, then its concentration in the blood increases and the patient becomes jaundiced. This type of jaundice is called *hæmolytic*.

Finally, if the liver itself is diseased the two pictures described above may be superimposed. Swelling of the liver cells interferes with the flow of bile through the small bile canaliculi and reabsorption into the blood follows, just as in obstructive jaundice. Since in this condition many of the liver cells are incapable of functioning at their normal level, bile pigments from the reticulo-endothelial system accumulate and a form of hæmolytic jaundice is also present. Such a condition is known as *toxic* or *hepatic jaundice*.

An attempt has been made to distinguish between the various types of jaundice by the *Van den Bergh test* in which the patient's serum is treated with Ehrlich's diazobenzene sulphonate reagent. It was at one time believed that two forms of bilirubin existed, one which had passed through the polygonal cells of the liver and gave an 'immediate direct' Van den Bergh reaction, and a second form which had not yet passed through the liver cells and which gave a 'delayed' reaction. A mixture of the two forms was supposed to give the 'biphasic' reaction. There is, however, no chemical evidence for the existence of two forms of bilirubin and the difference between the immediate and delayed reactions must be explained by the presence of interfering factors. Hence the Van den Bergh reaction is much more difficult to interpret than might at first sight appear and the test is not so commonly employed now as formerly.

When blood serum is treated with alcohol to precipitate protein and the diazo reaction applied to the protein-free filtrate, the 'indirect' Van den Bergh reaction is elicited. This test is very delicate and can be used for estimating bilirubin quantitatively in the serum, normal values lying between 0·2 and 0·8 mg. per 100 ml. A patient, however does not as a rule appear to be jaundiced until the serum bilirubin is about 2 mg. per 100 ml. A simpler method for assessing the degree of jaundice (hyperbilirubinæmia) is that of Meulengracht who compared the colour of diluted serum with that of a standard solution of potassium dichromate in order to obtain what is called the *icterus index*, normally 4 to 7.

REFERENCES

Drabkin, D. L. (1951). Metabolism of the hemin chromoproteins. *Physiol. Rev.* **31,** 345–363.

Gray, C. H. (1953). *The Bile Pigments.* London: Methuen.

Lemberg, R., & Legge, J. W. (1950). *Hœmatin Compounds and Bile Pigments ; their Constitution, Metabolism, and Function.* New York: Interscience.

Rimington, C. (1951). Pigments of blood and bile. *Lancet,* **2,** 551–553.

Rimington, C. (1952). Hæms and porphyrins in health and disease. *Acta med. scand.* **143,** 161–176. and 177–191.

Roughton, F. J. W., & Kendrew, J. C. (Editors) (1949). *Hœmoglobin.* London: Butterworth.

Williams, R. T. (Editor) (1954). The chemical pathology of animal pigments. *Biochem. Soc. Symp. No.* 12. London: Cambridge University Press.

Wolstenholm, G. E. W. & Millar, E. C. P. (Editors) (1955). Ciba foundation symposium on porphyrin biosynthesis and metabolism. London: Churchill.

CHAPTER 26

SUMMARY OF LIVER FUNCTIONS

SO many of the activities of the body are influenced by the liver that the detailed descriptions of its functions are necessarily scattered over many chapters. It is convenient here to present in summarized form an account of the structure and functions of the liver. The microscopic structure of the organ is described on p. 529; it consists mainly of polygonal cells (the liver parenchyma) together with portal tracts in which are conveyed the small bile ducts and blood vessels. From the chemical point of view the nucleus of the cell consists of deoxyribonucleic acid (together with a little ribonucleic acid which is concentrated mainly in the nucleolus), histone and several other more complex proteins. The complex of nucleic acid and histone is referred to by histologists as 'chromatin.' The nucleus also contains small amounts of lipoid material and a number of enzymes, including phosphatase.

The cytoplasm of the liver cell contains microscopic and submicroscopic granules. The granules visible in the microscope include particulate glycogen (p. 293) and basophilic particles 0·5 to 2·0 μ in diameter such as the mitochondria, which are complexes (p. 67) containing phospholipid, ribonucleic acid and protein. The submicroscopic particles, 50 to 200 mμ in diameter, are also phospholipid-ribonucleoprotein complexes and form the chromophilic ground substance of the cell.

The main functions of the liver may be summarized as follows:—

(1) **Carbohydrate Metabolism.**—As well as acting as a storehouse of glycogen, the liver is the site of the production of carbohydrate in the process of gluconeogenesis (p. 297).

(2) **Fat Metabolism.**—The liver is the site of fatty-acid oxidation and ketone body formation (p. 330).

(3) **Protein Formation.**—Synthesis of fibrinogen, prothrombin, and probably of plasma albumin also, occurs in the liver (p. 373).

(4) **Protein Metabolism.**—Deamination of amino-acids occurs in the liver and so also does urea formation (p. 342).

(5) **Nucleoprotein Metabolism.**—The formation of uric acid (p. 355) takes place in the liver.

(6) **Steroid Metabolism.**—Certain of the steroid hormones, in particular the œstrogens, are broken down in the liver (Chapter 52).

(7) **Blood-cell Formation and Breakdown.**—In the fœtus the liver is an important site of hæmopoiesis. In the adult the reticulo-endothelial cells lining the liver sinusoids play a part in hæmoglobin breakdown by forming the bile pigments which are later excreted into the bile by the polygonal cells (p. 414). The hæmopoietic factor (p. 132 and p. 392) is stored in the liver.

(8) **Bile Salts.**—The polygonal cells of the liver secrete bile salts into the bile as well as alkaline phosphatase and a number of foreign substances, for example the sulphonamides, tetraiodo-phenolphthalein and certain dyes.

(9) **Vitamins.**—The liver acts as a storehouse of vitamin A. This function is particularly well developed in certain fishes, for example cod,

halibut and tunny. Under optimal nutritional conditions the liver contains relatively large amounts of riboflavin and nicotinic acid in their combined forms (p. 125), and of ascorbic acid, vitamin D_3 and tocopherol.

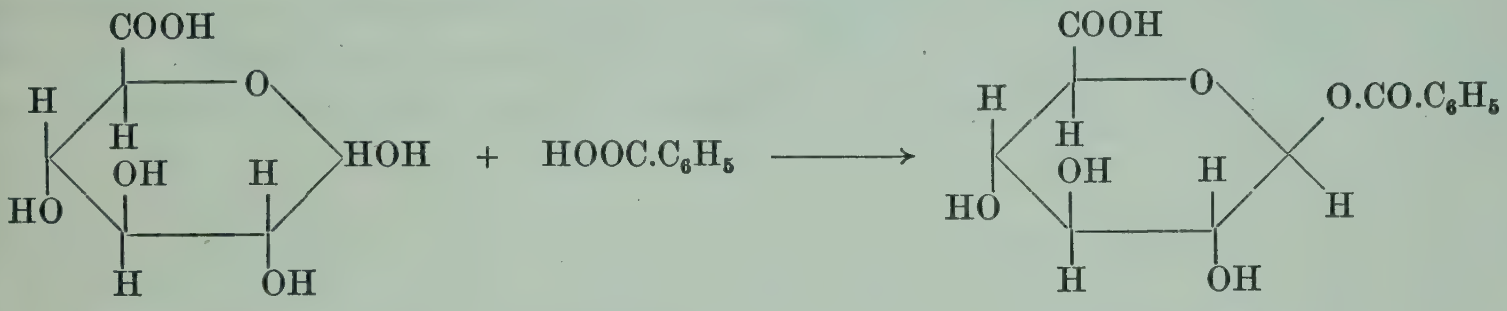
Since the liver occupies a key position in nutrition, disease of the liver implies a grossly disordered metabolism.

(10) ‘ Detoxication ’ Mechanisms:

(a) *Conjugation.*—An important function of the liver, which has not yet been dealt with, is conjugation, that is, its ability to couple foreign materials, many of which are toxic, with substances such as glycine or glucuronic acid to form relatively harmless complexes which are then excreted in the urine. For instance benzoic acid administered to man is conjugated with glycine in the liver and excreted as hippuric acid (p. 347):

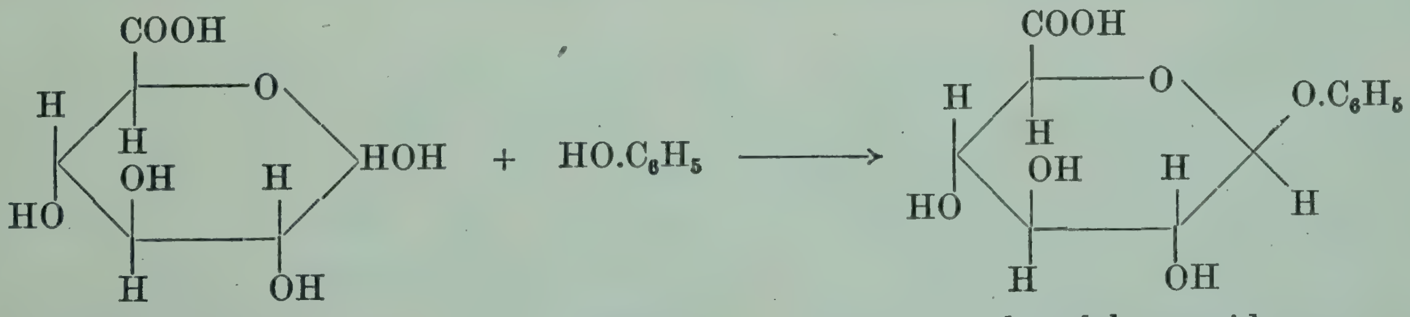
$$\underset{\text{benzoic acid}}{C_6H_5COOH} + \underset{\text{glycine}}{H_2N.CH_2.COOH} \longrightarrow \underset{\text{hippuric acid}}{C_6H_5CO.NH.CH_2.COOH} + H_2O$$

Benzoic acid may also be conjugated with glucuronic acid to give a glucuronide of the ester type (benzoylglucuronide):



glucuronic acid benzoic acid benzoylglucuronide

On the other hand phenol, which may be formed as the result of putrefactive breakdown of phenylalanine and tyrosine, is detoxicated by conjugation with glucuronic acid to give a glucuronide of the ether type (phenyl glucuronic acid):

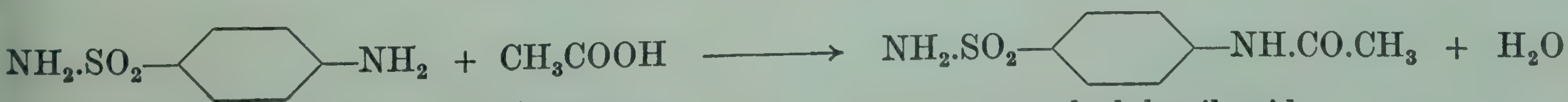


glucuronic acid phenol phenylglucuronide

Phenol may also be detoxicated by conjugation with sulphuric acid to form an organic derivative, the so-called ‘ethereal sulphate’ (p. 354):

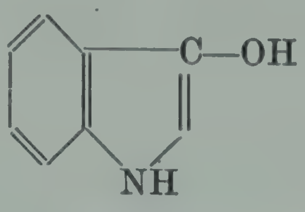
$$\underset{\text{phenol}}{C_6H_5OH} + \underset{\text{potassium hydrogen sulphate}}{HO.SO_2.OK} \longrightarrow \underset{\text{potassium phenylsulphate}}{C_6H_5O.SO_2OK} + H_2O$$

Acetic acid may also be involved in detoxication thus:

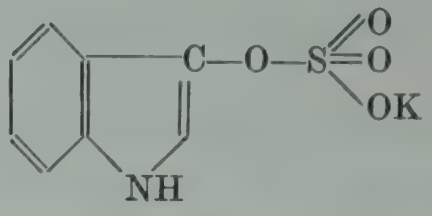


sulphanilamide acetic acid acetylsulphanilamide

Of the ethereal sulphates quantitatively the most important is that of indoxyl, derived from tryptophan. Indoxyl is normally excreted in the urine in amounts up to 20 mg. daily in the form of potassium indoxyl sulphate or *indican.*



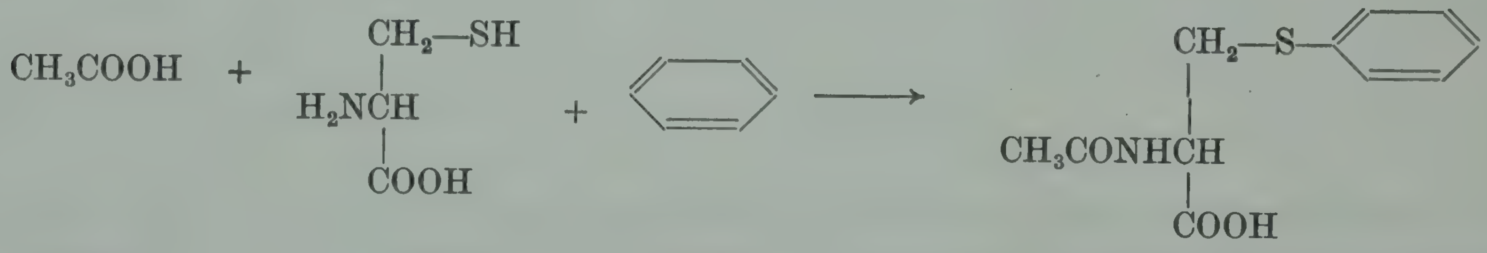
indoxyl



indican

The clinical significance of indicanuria is not clear. An increased urinary output of indican occurs, however, when the diet is rich in protein (and therefore in tryptophan). It is also found when large inflammatory exudates (such as abscesses) are being absorbed and when there is a rapid loss of tissue protein. It will be understood from what has already been said that the excretion of indican is dependent not only on its excess production and absorption but also upon adequate liver function. Tests based upon estimations either of indican or of ethereal sulphate in the urine have not, however, proved useful as liver function tests.

Some substances such as benzene, bromobenzene and naphthalene are coupled with cysteine, which in turn becomes acetylated, to form *mercapturic acids.* Benzene for example forms phenylmercapturic acid thus:—

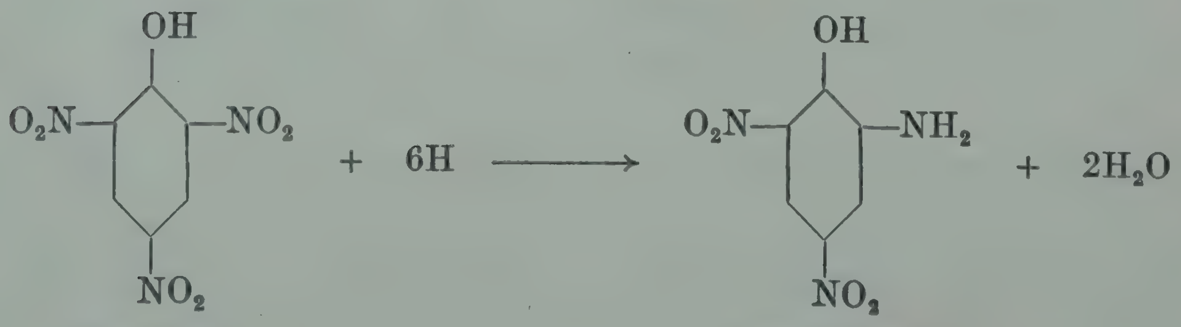


acetic acid cysteine benzene phenylmercapturic acid

Such mercapturic acids are found in the 'neutral sulphur' fraction of the urine (p. 354).

(b) *Oxidation and Reduction.*—The body cannot readily oxidize toxic aromatic compounds and they are for the most part dealt with by conjugation. Aliphatic, or open chain, compounds are more easily oxidized and small amounts of cyanide, for example, are said to be detoxicated by oxidation in the liver to non-toxic thiocyanates, which are then excreted through the kidneys and salivary glands. In a number of cases a toxic substance is converted to a less harmful one by reduction.

$$\underset{\text{chloral}}{CCl_3.CHO} + 2H \longrightarrow \underset{\text{trichloroethyl alcohol}}{CCl_3.CH_2OH}$$



picric acid picramic acid

HEPATIC FUNCTION TESTS

Increasing knowledge of the many processes carried out by the liver has led to the development of numerous chemical tests for assessing the efficiency of hepatic function in disease. It is not possible to mention here all the procedures that have from time to time been devised for this purpose or to describe in detail any single test. In broad terms, clinical tests of liver function can be divided into three groups:

GROUP I. TESTS BASED UPON SYNTHESIS OF PROTEIN BY THE POLYGONAL CELLS.—Acute inflammatory processes in the liver may be associated with qualitative alterations in the proteins of the blood serum. The addition of such a serum to certain colloidal solutions (gold salts, thymol, cephalin-cholesterol) results in precipitation which is not produced by normal serum. In chronic liver disease the concentration of the serum albumin tends to fall; in addition the concentration of prothrombin in the plasma may be reduced and the increase which should follow the administration of vitamin K analogue (p. 148) may be either absent or inconspicuous.

GROUP II. TESTS BASED UPON THE METABOLIC FUNCTION OF THE LIVER.—Extensive or long-standing disease of the liver parenchyma is often associated with evidence of disordered hepatic metabolism. For example the synthesis of hippuric acid (see p. 347) may be defective, or it may be found that the ingestion of an amount of fructose or galactose, insufficient to increase greatly the blood sugar in a healthy individual, is followed by hyperglycæmia and glycosuria (p. 296).

GROUP III. TESTS OF EXCRETORY FUNCTION.—If liver function is impaired serum concentration of bilirubin (icteric index, Van den Bergh reaction) may be increased and bilirubin, or an abnormal amount of urobilin, may be detected in the urine (p. 416). An increased concentration of alkaline phosphatase in the serum, a raised serum cholesterol level, or the presence of an abnormal amount of dye (bromsulphalein) in the blood an hour after the intravenous injection of a standard dose of dye may likewise indicate impaired function.

In investigating hepatic function it is unwise to rely on any one test since a single abnormal result is not unequivocal evidence of hepatic disease. A good principle is therefore to employ a single test repeatedly or, better, to perform a number of tests, including if possible one from each of the three groups. The reserve capacity of the liver is such that many tests do not yield abnormal results until there is extensive, and perhaps irreparable, damage to the liver tissue.

REFERENCES

SHERLOCK, S. (1955). *Diseases of the liver and biliary system.* Oxford: Blackwell.

STEWART, C. P., & DUNLOP, D. M. (1954). *Clinical Chemistry in Practical Medicine*, 4th ed. Edinburgh: Livingstone.

WILLIAMS, R. T. (1947). *Detoxication Mechanisms.* London: Chapman & Hall.

CHAPTER 27

CIRCULATION

INTRODUCTION TO CIRCULATION

THE circulation of the blood through the body has so long been accepted as a fact of everyday life that the observations which originally led Harvey to its discovery are often forgotten. His experiments were first recorded in his book *De motu cordis*, published in 1628. Before Harvey's time all sorts of quaint ideas about the motion of the blood were current but many facts were accumulating which led Harvey inevitably to the correct conclusion. This in no way detracts from Harvey's achievement but it is nevertheless true that discoveries are made when the time is ripe for them. Harvey's work is an illustration of the modern method of investigation in which the results of observation form the basis of a working hypothesis which, if supported by sufficient experimental findings, becomes an accepted theory. Harvey's main observations are given here.

When the heart is held in the hand it becomes harder and smaller as it contracts just as does a muscle of the arm. The contraction of the heart occurs a very short time before the expansion of the arteries suggesting that the contraction of the heart is responsible for the arterial pulse. When an artery is cut the blood gushes out from the end nearer the heart; the force of the gush is increased at each contraction of the ventricles so that the blood appears to leave the artery in spurts. Once the blood has passed into either of the great arteries it cannot flow backwards into the heart because of the semilunar valves (as Leonardo da Vinci had demonstrated). The flow into the arteries must, therefore, always be from the heart to the periphery. A simple calculation based on this drives us inevitably to the idea of a circulation. Suppose the ventricle ejects 60 ml. at each beat and there are 72 beats per minute. In one hour the heart will eject $60 \times 72 \times 60 = 259{,}200$ ml. $= 259{\cdot}2$ litres. This is more than three times the volume of a man and very many times the volume of his blood. There is no possible mechanism for supplying this large volume of fluid every hour —the only way such a large flow could be maintained is by some form of circulation. Further, when the vena cava is clamped, the vein between the clamp and the heart collapses and at the same time the heart becomes empty and pale. When the aorta is clamped the heart and the aorta up to the clamp are distended. If a bandage is put round the arm tight enough to obstruct the superficial veins but not enough to obstruct the arteries the veins swell up. This blood cannot come from the veins; it must come from the arteries. Moreover the veins contain valves which all look towards the heart and prevent flow in the direction away from the heart. This can easily be shown in the forearm (Fig. 27, 1). When a moderately tight bandage is placed round the arm the veins fill up and the positions of the valves are recognized by slight dilatations at short intervals along the vein. If an attempt is made to massage the blood away from the heart the valves stand out more prominently but the blood cannot be impelled onwards unless the

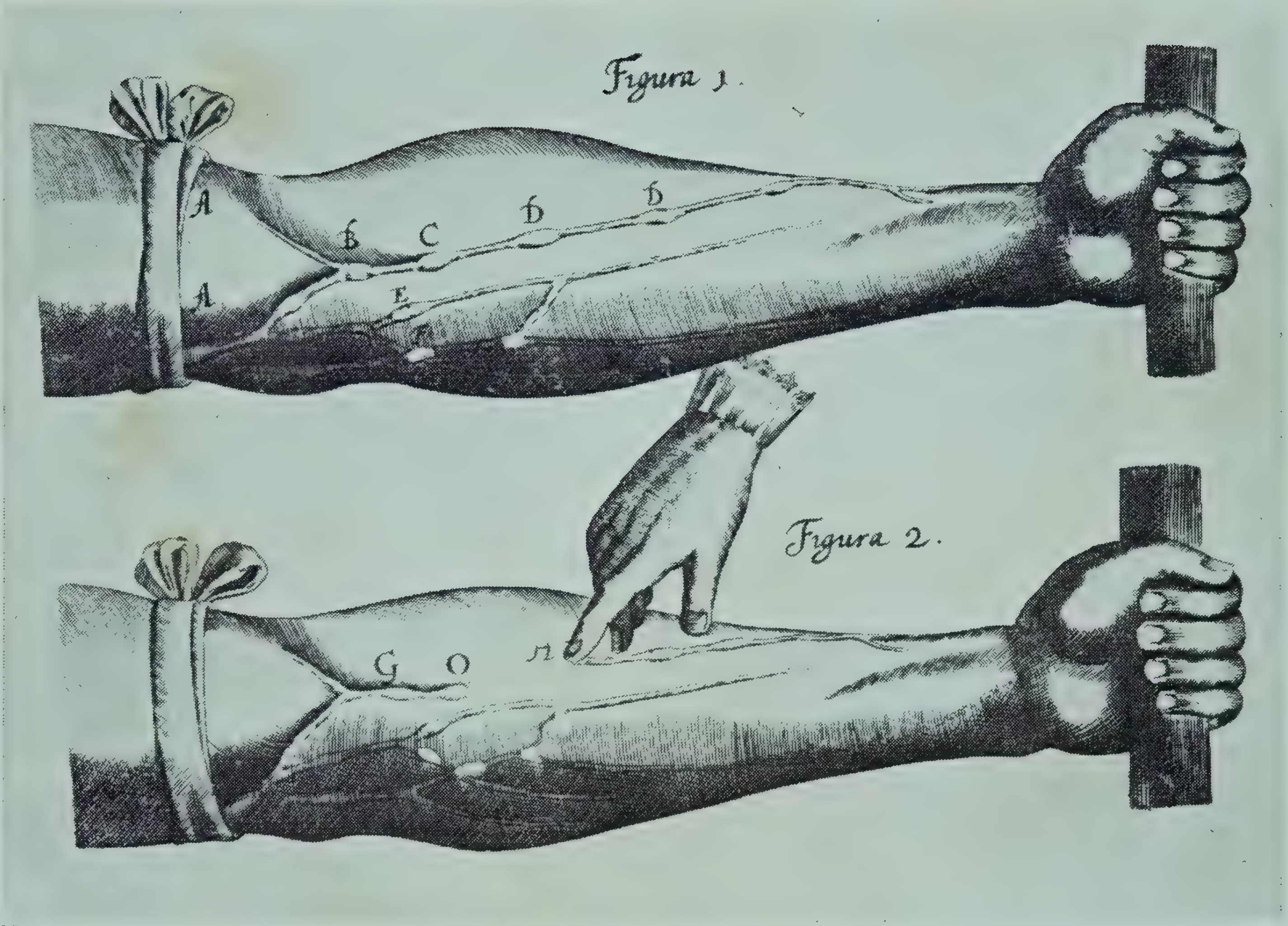


Fig. 27, 1

Illustrations from Harvey's "De motu cordis" (1628) (Frankfurt) to illustrate the valves, A, B, C, D, in the veins. In the lower diagram a vein has been compressed by the finger at *r*, the blood in section *r*O has been massaged towards the heart. Flow in the direction GO is prevented by the valve at O.

valves are forcibly ruptured. Further if a portion of distended vein is compressed distally by the finger it may be emptied by massaging it in a proximal direction towards the next valve (Fig. 27, 1). The vein does not fill again until the distal finger is removed, indicating that the flow of blood in the vein proceeds from the periphery towards the centre. Harvey also noticed the difference in the thickness of the walls of the arteries and veins; the greater thickness of the arteries was, he supposed, to sustain the shock of the impelling heart and to accommodate the blood that was expelled from it.

Harvey could only have a vague notion of the manner of the flow of the blood between arteries and veins through the 'pores of the flesh.' Malpighi of Bologna about 1661 was the first to see the circulation through the capillaries in the lung of the frog. He described how the blood is forced from the arteries into a network of capillaries and then passes into the receiving branches of the veins.

The circulation of the blood is particularly easily observed in the vessels of the mesentery, where it may be seen that the blood in a small artery is flowing in one direction, whereas in the companion vein it flows in the opposite direction.

A plan of the circulation in man is shown in Fig. 27, 2.

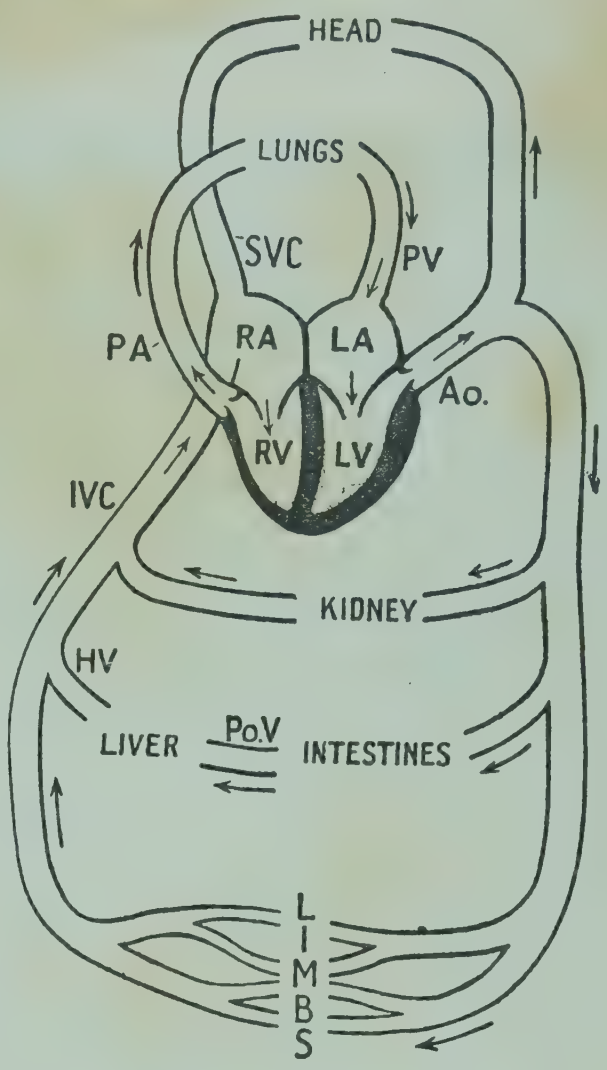


Fig. 27, 2

General plan of the circulation. PA, pulmonary artery. PV, pulmonary vein. LA, left atrium. LV, left ventricle. Ao, aorta. Po.V, portal vein HV, hepatic vein. IVC, inferior vena cava. SVC, superior vena cava. RA, right atrium. RV, right ventricle.

In all but the simplest animals a circulation of body fluid is necessary for a large number of reasons. The most urgent need of the tissues, especially the contractile tissues, is for oxygen. The activity and efficiency of the animal depend to a large extent on the ability of the circulation to provide an adequate supply of oxygen. In addition the tissues obtain their foodstuffs from the circulating blood and pass their waste products into it.

In the following chapters we shall deal with the heart and blood vessels and the regulation of their activities, and later with the respiratory function of the blood.

REFERENCES

McDowall, R. J. S. (1938). *The Control of the Circulation of the Blood.* London: Longmans.
Singer, C. (1922). *The Discovery of the Circulation of the Blood.* London: G. Bell & Sons.

CHAPTER 28

THE HEART

THE first rudiment of the human heart is a simple tube, formed by the fusion of two aortæ, which later develops dilatations corresponding to the chambers of the primitive heart. By analogy with the chick heart, which can be readily observed *in vivo*, there is no doubt that the earliest movement must be a kind of peristalsis which forces the blood from the sinus to the atrium and on to the ventricle, bulbus and truncus. At a very early stage the heart becomes **S**-shaped and then by a series of complicated changes it reaches the adult form. The primitive sinus venosus, where the contraction of the embryonic heart begins, is represented in the adult by two small nodes of tissue, one at the termination of the superior vena cava (sinu-atrial node) and the other near the mouth of the coronary sinus in the right atrium (atrio-ventricular node, Fig. 28, 27).

ANATOMY OF THE HEART

The adult human heart weighs from 300 to 350 g. It possesses two upper chambers, the *atria* (or auricles in the older terminology), and two lower chambers, the *ventricles*. As the blood is held up in the atria for only a part of each cardiac cycle their capacity is about two-thirds that of the ventricles. During the remainder of the cycle the blood passes freely from the large veins into the ventricles. The walls of the atria are thin, corresponding with the small amount of work which is required to expel their contents into the ventricles. The ventricles do most of the work of the heart, pumping the blood out at high pressure into a system of branching tubes which offer considerable resistance to the onflow of the blood, and accordingly the walls of the ventricles are much thicker than those of the atria. The wall of the left ventricle one inch below the atrio-ventricular ring is about half an inch thick. This is three or four times thicker than the wall of the right ventricle at the corresponding position and is related to the greater amount of work required of the left ventricle (p. 447).

The volume of the human heart during life is some 700 ml. at the end of a contraction whereas the actual volume of muscle is about 300 ml. Thus the cavities must contain roughly 400 ml. of blood, an amount that is much greater than the quantity expelled by both ventricles each time they contract (about 120 ml.). Presumably therefore the cavities are not completely empty at any time. This conclusion is supported by X-ray studies made while the heart contains a contrast medium introduced rapidly into the venous blood stream (angio-cardiography). Such photographs show that at the end of systole all the chambers of the heart, except the left ventricle, contain blood.

Each ventricle has the same capacity and each expels the same amount of blood at systole or contraction. If this were not so the circulation would eventually cease. Let us consider what would happen if the right

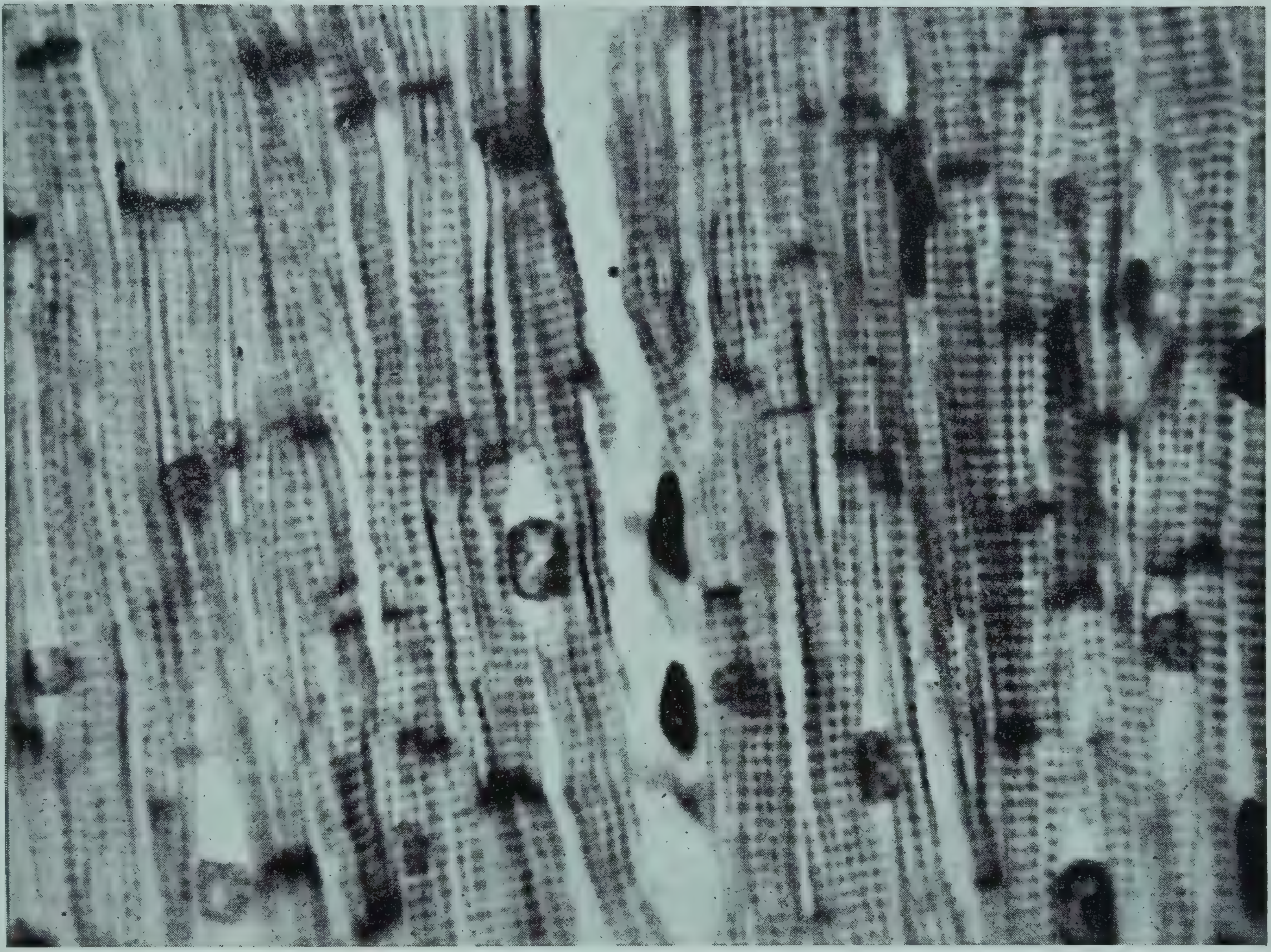


FIG. 28, 1

Photomicrograph of human heart muscle-cells (magnification 900 times). Longitudinal and transverse markings are well shown. The dark transverse markings are intercalated discs.

(By courtesy of H. S. D. Garven.)

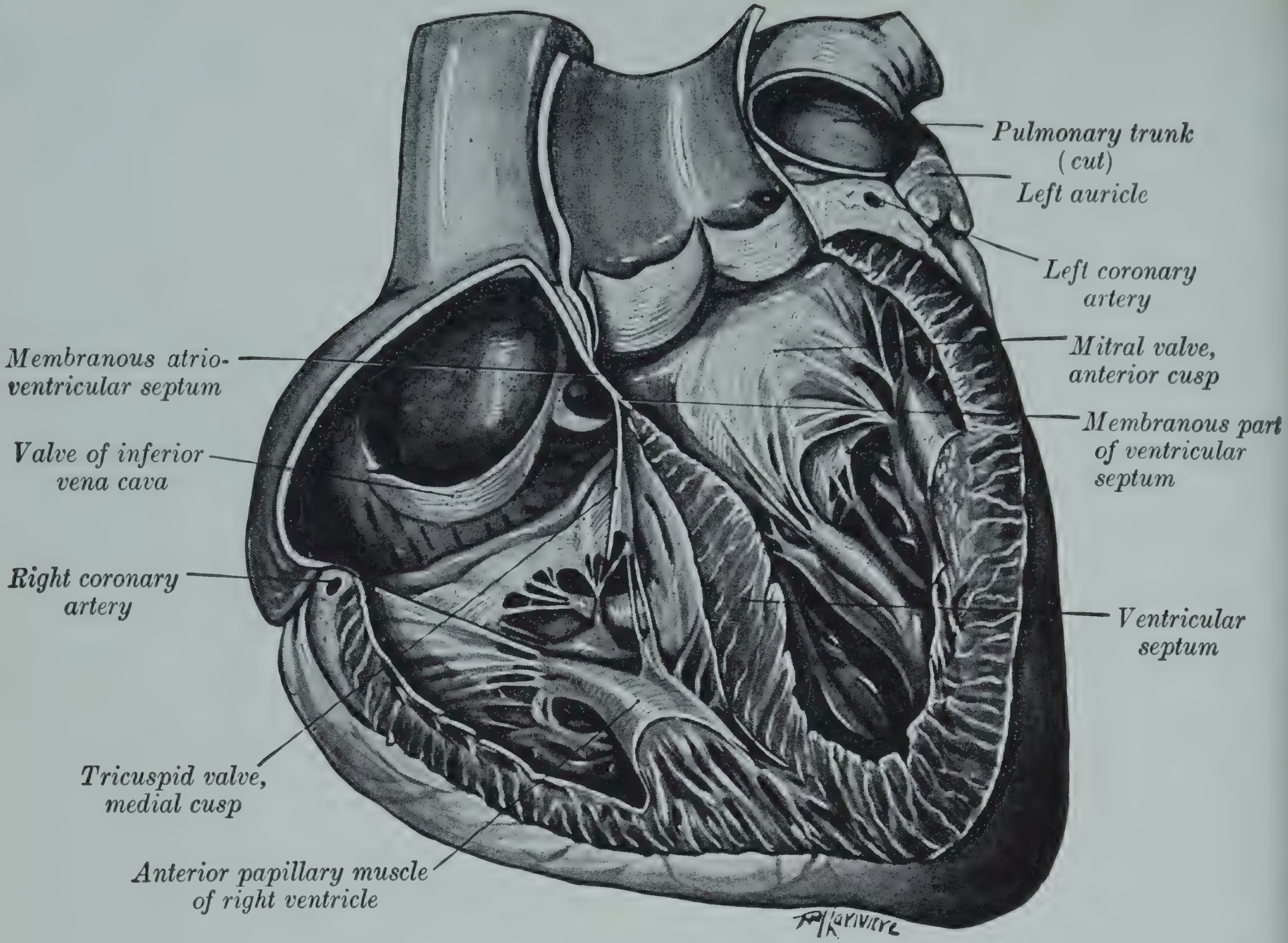

Pulmonary trunk (cut)
Left auricle
Left coronary artery
Mitral valve, anterior cusp
Membranous part of ventricular septum
Ventricular septum
Membranous atrio-ventricular septum
Valve of inferior vena cava
Right coronary artery
Tricuspid valve, medial cusp
Anterior papillary muscle of right ventricle

FIG. 28, 2

Section through the human heart to show the valves. On the right side the atrio-ventricular orifice has been cut across. On the left side the section passes through the aortic orifice and in front of the atrio-ventricular orifice (*Gray's Anatomy*, 31st ed. (1954), p. 694).

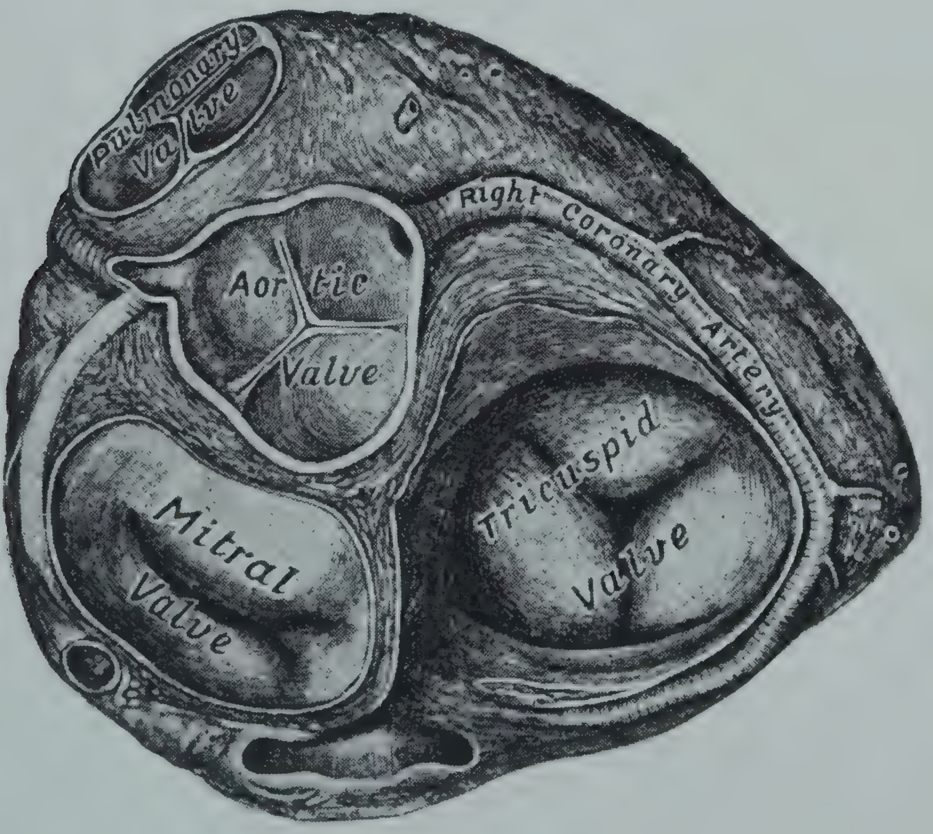
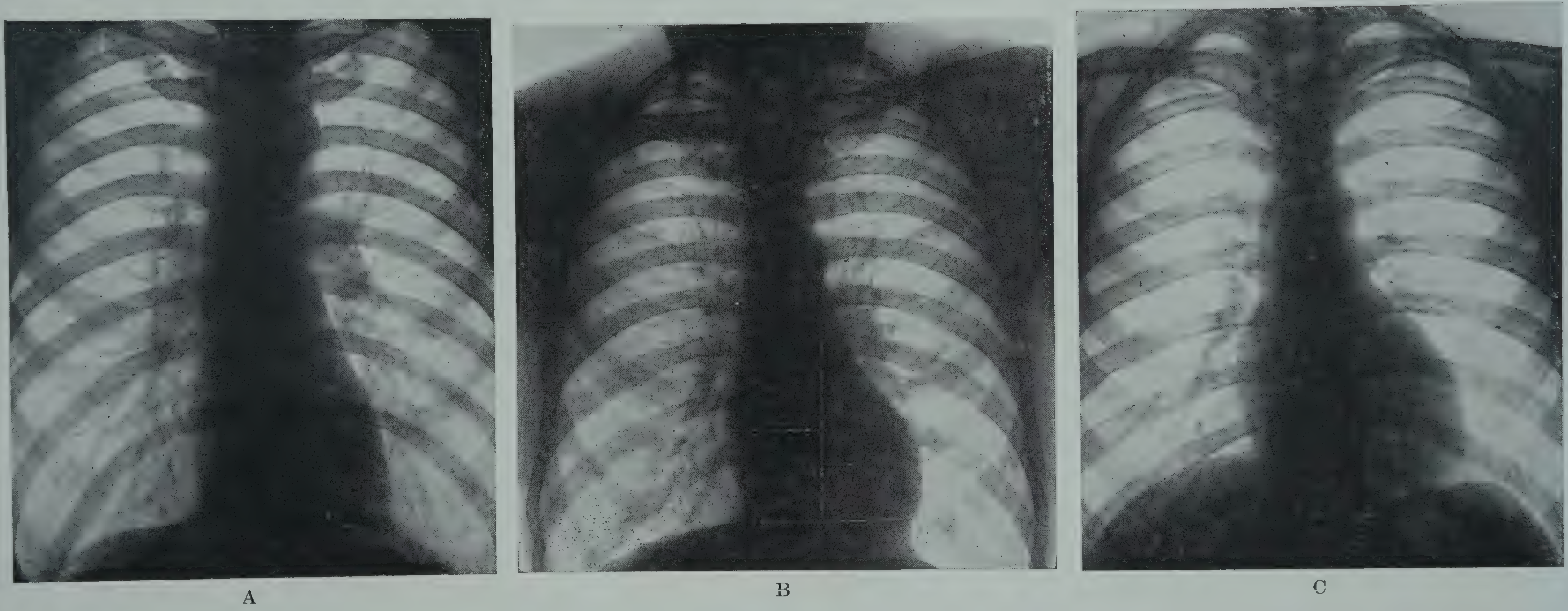

Pulmonary Valve
Aortic Valve
Right Coronary Artery
Mitral Valve
Tricuspid Valve

FIG. 28, 3

The bases of the ventricles have been exposed by removal of the atria to show the heart valves (*Gray's Anatomy*, 22nd ed. (1923), p. 585).



A

B

C

FIG. 28, 4

Teleradiograms of the chests of three healthy men of varying body build to show the extent of variation of the normal cardiac shadow. The subject in FIG. 28, 6, A was tall and slender, while the subject of FIG. 28, 6, C, was short and stocky. All the photographs were taken during quiet respiration with the subject standing.

ventricle were to deliver, say, 0·1 ml. more blood per contraction than the left. If the left expels 60 ml. per beat and the right 60·1 ml. per beat then in one hour, if each beats 80 times a minute, the left expels 288,000 ml. and the right 288,480 ml. The difference, 480 ml., would have to be accommodated in the pulmonary vessels and lungs, with grave effects upon the pulmonary circulation and the vital capacity. This does not mean, however, that for *short* periods of time the amounts of blood expelled by the two ventricles must be exactly the same; at certain phases of the respiratory cycle, for example, there are small differences between the right and left stroke volumes. The average outputs must, as just explained, be equal.

The atria and the ventricles are completely separated by a *fibro-tendinous ring* which gives attachment to most of the muscular fibres and also to the atrio-ventricular valves. The only muscular connexion between atria and ventricles is at the relatively small A.V. node. Closely connected with the main fibrous ring are other fibrous rings which surround the arterial orifices and serve for the attachment of the great vessels and semilunar valves and some of the ventricular muscle fibres. While the superficial muscle fibres pass transversely round both atria, the deeper fibres loop over from the front to the back of each atrio-ventricular ring, some forming rings round the openings of the great veins into the heart. In the ventricles the muscle fibres are arranged in a very complicated fashion. Superficial fibres arise from the fibrous rings, sweep down spirally to the apex and end in the papillary muscles which project into the ventricular cavity; deep fibres arise in the papillary muscles of one ventricle and pass in an S-shaped manner through the interventricular septum to the papillary muscles of the other ventricle.

The cells of the cardiac muscle are brick-shaped or irregularly four-sided as seen in a histological section (Fig. 28, 1). Occasionally one or both ends of a cell have short branches by which adjacent cells are joined. Like skeletal muscle, cardiac muscle shows both longitudinal and transverse striations, but the cardiac muscle cell is characterized by a central nucleus, by the poorly developed sarcolemma and by the fact that the contractile cells are not separate units but form a syncytium with protoplasmic, and consequently functional, continuity between the units. The cells of the *sinu-atrial* (S.A.) *node* and *atrio-ventricular* (A.V.) *node* are narrow and short and have interlacing branches. From the A.V. node a bundle of special conducting tissue composed of modified muscle cells called the A.V. *bundle* passes down to the ventricles. These cells are broad, with a faint peripheral striation, and have a high glycogen content. The A.V. bundle divides into right and left fasciculi which pass to the endocardial surface of the ventricles and join the sub-endocardial plexus of Purkinje fibres. These blunt branching cells arranged end to end are modified cardiac muscle cells.

Heart Valves.—The heart valves (Fig. 28, 2 and 3) are formed by flaps of fibrous tissue attached to the fibrous rings and covered on both sides by endocardium. All the valves are so constructed that they permit the blood to flow only from the atria to the ventricles and onward into the arteries. The points of entry of the veins into the atria are not provided with valves except for the inferior vena cava which has a rudimentary valve, but the arrangement of the loops of atrial muscle round the orifices of the veins may tend to prevent reflux of blood. Back-flow of blood is known to take place, however, in the coronary sinus in man.

The *atrio-ventricular valves* are large, sail-like structures which arise from the atrio-ventricular rings and pass down into the cavity of the ventricles (Fig. 28, 2). Their margins are connected by thin cords (*chordæ tendineæ*) to the papillary muscles which form part of the muscular mass of the ventricles. As the ventricles contract the papillary muscles pull on the chordæ tendineæ so that the flaps of the valves cannot be turned back into the atria. On the left side, that is to say between the left atrium and left ventricle, there are two flaps forming the mitral valve; on the right side the tricuspid valve has three flaps.

The arterial orifices are guarded by the *semilunar valves*. When the aorta, or pulmonary artery, is cut open three semilunar flaps or cusps of very thin fibrous tissue covered by endocardium are exposed. In the middle of the free edge of each valve cusp there is a little thickening. In the intact heart the three pockets can meet along three lines and prevent blood from flowing back into the ventricle (Fig. 28, 3).

Coronary Vessels.—The heart muscle is supplied with blood by the right and left coronary arteries which arise from the aorta in two of the three pockets (aortic sinuses) lying behind the cusps of the semilunar aortic valve (Fig. 28, 3). The venous blood is collected chiefly into the coronary sinus which opens into the right atrium; some small veins open independently directly into the right atrium and numerous smaller veins (Thebesian vessels) drain blood directly into the cavities of the ventricles, especially the right.

The Pericardium.—The heart is enclosed in a conical sac, the fibrous pericardium, the neck of which merges with the fibrous covering of the walls of the great vessels at the base of the heart; the base of the sac is attached to the central tendon of the diaphragm. Its inner surface and the outer surface of the heart are lined by a single layer of flattened cells resting on a layer of loose connective tissue. This is the serous pericardium. The two serous surfaces are separated by a thin film of fluid which allows free movement of the layers over one another.

The Size of the Heart

In the adult male the heart is about 0·43 per cent. of the body weight and in the female about 0·4 per cent. but in active animals like the deer or greyhound the heart is about 1 per cent. of the body weight. Although at the bedside attempts are made to estimate the size of the heart by the clinical methods of palpation and percussion, exact knowledge of cardiac size can be derived only by X-rays. Since the X-ray picture is really the *shadow* of the heart it has to be remembered that the size of a shadow depends on the distance of the light source (in this case the X-ray tube) from the object, and also upon the distance from the object, that is the heart, to the surface of the photographic film on which the shadow is cast. Photographs taken with a tube-to-film distance of two metres, at the end of a quiet expiration with the patient as close as possible to the film, give negligible distortion with a magnification of the cardiac area of some 2 to 10 per cent. Such photographs which are taken during diastole are called *teleradiograms*.

The shadow of the normal heart and blood vessels is remarkably variable, depending on the position of the subject and on weight, build and muscular activity. The three X-ray photographs of Fig. 28, 4, illustrate some variations in the shape of the heart shadow in normal persons. The shape of the heart as judged by X-rays alters as between deep inspiration and deep expiration, but there is usually little alteration with exercise. Elevation of the diaphragm from pregnancy or increased intra-abdominal pressure, however, modifies the shape of the cardiac shadow and alters the position occupied by the heart in the chest. The transverse diameter of the cardiac shadow averages 12·2 cm in the male and 11 cm in the female but there is a wide range of normality—from 8 to 14·5 cm according to Roesler. If a teleradiogram is taken of the heart in the lateral position the antero-posterior diameter of the shadow can be measured. The mean of many such measurements is 9 cm with a range of 7 to 11 cm. One of the most useful indices of cardiac size is the cardio-thoracic ratio, that is the transverse diameter of the heart shadow divided by the maximal transverse diameter of the shadow of the chest; this should not exceed 0·5 but in children up to four years of age this ratio is relatively greater than it is in adults. Much information about even minor enlargements of the various chambers of the heart can be obtained by detailed X-ray examination (cardioscopy).

THE FROG'S HEART

The student of physiology frequently uses the frog's heart as experimental material. Since many of the properties of cardiac muscle are very easily demonstrated by the use of this preparation a short description of the anatomy and physiology of the frog's heart is necessary.

In the frog the venous blood reaches the sinus venosus by three venæ cavæ and then passes into the right atrium; the pulmonary veins drain into the left atrium. Both atria communicate with the single ventricle which pumps the mixed arterial and venous blood through the truncus to the two aortæ. Since there is no definite fibrous ring between atria and ventricle there is complete muscular continuity throughout the heart, and there is accordingly no special conducting tissue comparable to the A.V. bundle of the mammalian heart. The fibres of the vagus nerve go to Remak's and other ganglia in the heart and postganglionic relay fibres carry the impulses to the muscle cells. The sympathetic fibres leave the spinal cord in the third spinal nerve and, passing through the ganglion of the vagus, run with the vagal fibres as the vagosympathetic nerve which enters the heart at the junction between sinus venosus and atrium (the crescent).

Rhythmic Activity.—The frog heart beats regularly in the order sinus, atria, ventricle and truncus. If a ligature is tied round the junction of sinus and atria (first Stannius ligature) the sinus continues to beat but the remainder of the heart stops in a relaxed condition for some time. Atrial and ventricular contractions eventually recur but at a slower rate than the original sinus rate. If another ligature (second Stannius ligature) is tied round the heart between the atria and the ventricle, soon after the first ligature is applied, then the ventricle begins to beat again at quite a slow rate. In other words the natural rate of working of the various parts of the heart diminishes in the descending order—sinus, atrium, ventricle and truncus. These differences in inherent rate can also be observed by separating the heart into its four parts which are then placed in Ringer's solution (p. 431). The muscle tissue from the sinus beats faster than that from the atrium and the atrial rate is higher than the ventricular.

At the end of systole the ventricle relaxes and it would, in the absence of impulses reaching it from the atria, contract again spontaneously after a certain relatively long interval. In a normal cycle, however, the sinus contracts and an impulse passes to the atria and on to the ventricle, which in this way is made to contract much earlier than it would have done if it had been left to itself. Because it possesses a higher frequency of rhythm than the other chambers, the sinus determines the rate of the heart and it is therefore called the *pacemaker*.

It is easy to show that the most rapidly acting part of the frog heart controls its rate. If the ventricle is stimulated electrically to contract at a rate faster than the sinus rate then the chambers contract in the reverse order—namely, ventricle, atria, sinus. Further, if the sinus, the normal pacemaker, is warmed then the heart rate increases but if the ventricle is warmed the rate is not affected.

The rhythmic activity of the heart does not depend on nervous reflexes but is a fundamental characteristic of the muscle. The best proof of this is given by tissue-culture experiments in which heart muscle cells show rhythmic movements in the entire absence of nerve elements. Two cultures of chicks' hearts grown close together beat quite independently until the two cultures merge with one another, when the whole culture pulsates synchronously. The actual chemical or physical origin of the rhythm is still a matter for speculation.

In the frog the excitatory wave is carried quickly over the atria and then over the ventricle by means of protoplasmic connexions between the cells and not by way of nerve fibres or specialized muscle tissue. The slight pause between the contraction of the atria and that of the ventricle allows time for the blood to get from the atria into the ventricle. The pause may be due to a partial block to the passage of the excitatory wave at the junction of the atria and ventricle; a similar block can be produced artificially at any other part of the heart by gentle compression or by cutting away some of the muscle tissue. It has been pointed out that in the frog the muscle fibres at the atrio-ventricular junction run circularly round the heart and not directly from atria to ventricle. This arrangement could also account for the delay between atrial and ventricular contraction.

'All-or-nothing Law.'—If a frog's heart, brought to a standstill by the first Stannius ligature, is stimulated electrically through electrodes laid on the ventricle, it is found that the period of contraction is about 0·5 sec. Further, the force of the contraction is always maximal; a weak stimulus does not elicit any response whereas all stimuli above the threshold value produce contractions of the same force. Since all the fibres are interconnected a stimulus, if effective, makes all the fibres contract. The 'all-or-nothing' law holds for cardiac muscle as well as for skeletal muscle (see p. 781) and nerve (see p. 739).

Refractory Period.—The refractory period of skeletal muscle is very short—very much less than the contraction period—so that one contraction can be followed very quickly by another. In this way a sustained contraction or tetanus can be produced by a rapid series of stimuli. In cardiac muscle, however, the refractory period is very long, at least as long as the contraction period. Thus heart muscle cannot normally be tetanized since any stimulus occurring during the contraction period is ineffective, that is, it falls in the *absolutely refractory period.* For a short time after the contraction is over a greater stimulus than usual is necessary to induce another contraction—this is the *relatively refractory period.* These findings can be explained by assuming that, when the heart contracts, chemical substances are broken down which must be built up again to a certain concentration before the muscle becomes responsive to any further stimulation. The refractory period could accordingly vary with the extent of breaking down, and therefore with the amount of building up, required. In a rapidly beating heart the stimuli from the pacemaker are falling within the relatively refractory phase before the chemical restoration

TABLE 28, 5

The relation between the length of the absolutely refractory period of cardiac muscle of the dog and the cardiac rate (Lewis)

Rate per Minute	Length of Cardiac Cycle sec	Duration of Absolutely Refractory Phase sec
100	0·6	0·2
150	0·4	0·17
290	0·2	0·11

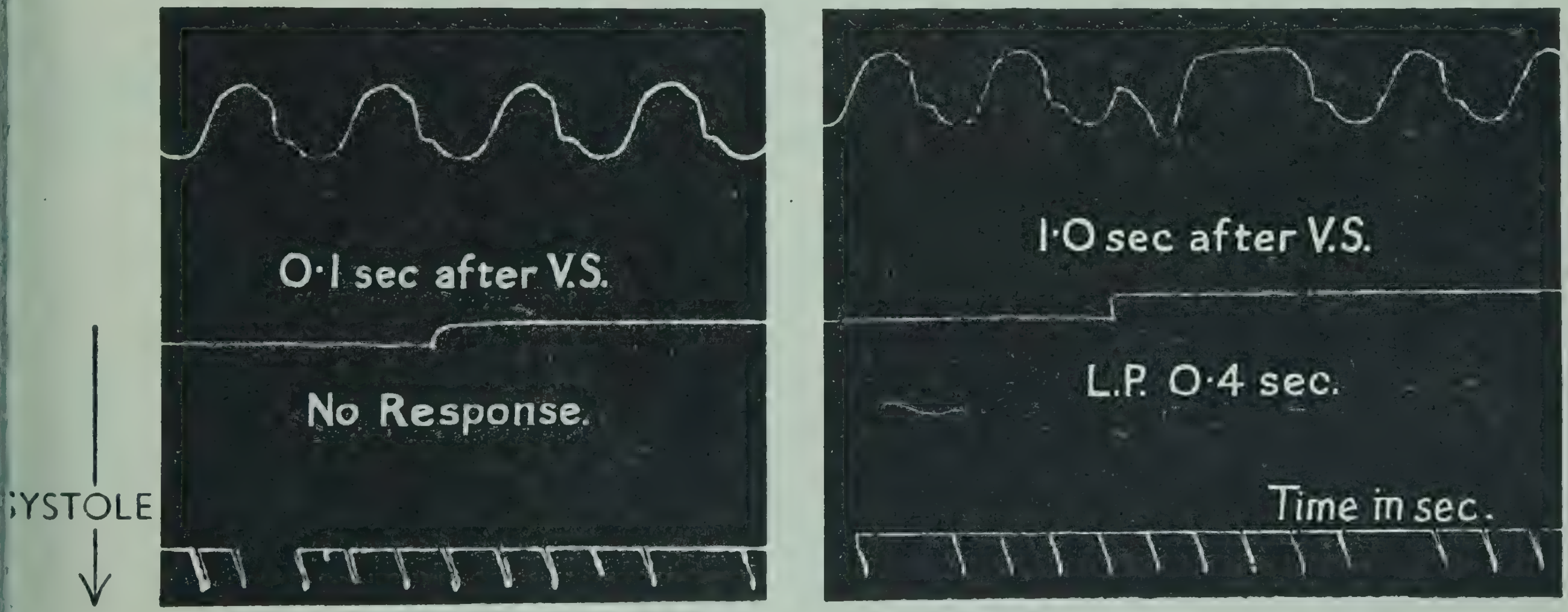


FIG. 28, 6

A. Stimulation at 0·1 sec after beginning of ventricular systole. No response. Stimulus falls within absolutely refractory period. Time in seconds. Middle trace signals stimulus.

B. Stimulation at 1·0 sec after beginning of ventricular systole. A premature contraction after a latent period of 0·4 sec completely replaces the next normal ventricular systole because the muscle is absolutely refractory when the stimulus to initiate this contraction arrives from the atria.

(*By courtesy of A. M. Halliday.*)

is complete. Smaller contractions therefore result because there is less material to provide energy, and the refractory period is shorter because there is less material to be built up. This conception is supported by Lewis's experiments on dog's atrial muscle in which he was able to show that the refractory period did actually diminish when the rate was increased (Table 28, 5). The biological advantage of the long refractory period, which makes a tetanus of the heart muscle impossible under natural conditions, is clear; a spasm or tetanus of the heart would bring the circulation to a standstill.

If electrodes are attached to the ventricle of a spontaneously beating frog's heart the ventricle can be stimulated electrically at different points of the cardiac cycle (Fig. 28, 6). When the ventricle is stimulated during its period of relaxation (*diastole*) a contraction (*systole*) is initiated before the next stimulus reaches it from the pacemaker. This premature contraction or *extrasystole* is followed by a longer pause than usual (sometimes called a compensatory pause), because the next spontaneous impulse travelling down from the sinus arrives at the ventricle during its absolutely refractory period. The ventricle remains in diastole until the second natural impulse arrives from the sinus, when the normal rhythm is resumed. The first contraction after the pause may be more forcible than normal because, as a result of the unusually long diastole preceding it, the heart contains more blood.

PERFUSION OF THE FROG'S HEART

The frog's heart can be perfused with fluid by tying a cannula into the posterior vena cava (Fig. 28, 7). The perfusion fluid must have the same osmotic pressure as that of the tissues, otherwise disruption of the muscle fibres would occur by shrinking or expansion of the cells. If a frog's heart is perfused with 0·7 per cent. NaCl (which is isotonic with frog's cells) the beats gradually diminish and finally cease. Contractions begin again when a trace of $CaCl_2$ is added to the perfusing fluid but the relaxations between

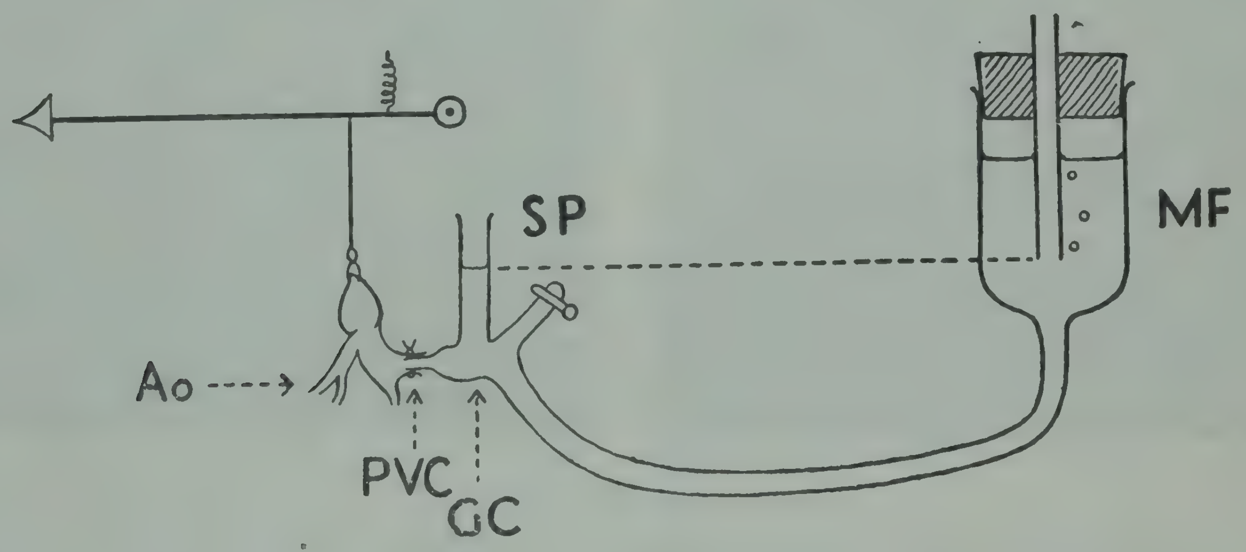


Fig. 28, 7

Diagram of the method of perfusion of the frog's heart by Greene's cannula (GC). The head of pressure is kept constant by using a Mariotte's flask (MF), containing Ringer's solution. The head of pressure in the atrium is read on the stand-pipe (SP). The fluid enters the heart through the posterior vena cava (PVC) and leaves through the aortæ (Ao).

beats become less and less and finally the heart stops in spasm, that is, tonically contracted. If now a trace of KCl is added, the contractions begin again and may go on for hours, although there is nothing in the fluid from which the muscle can derive the energy for contraction. However, cardiac contractions are not maintained when the perfusing fluid contains only $CaCl_2$ and KCl. The sodium ion is necessary to maintain irritability but the manner in which it does so is only recently beginning to be understood (see Chapter 41). Further experiments on these lines by Ringer in 1883 showed that the following is a suitable fluid for perfusion of the frog's heart.

Ringer's Solution

NaCl	0·6 g.
KCl	0·0075 g.
$CaCl_2$	0·01 g.
$NaHCO_3$	0·01 g.
Distilled water,	100 ml.

The heart shows maximum activity if the pH of the perfusion fluid is about 8. A frog's heart perfused with a slightly acid solution dilates and very acid (pH 3) solutions make the beats very weak, the heart stopping in diastole. With alkaline solutions relaxation becomes less complete and the heart finally stops in systole. If the fluid perfusing the frog's heart is warmed the heart quickens; at the same time the conduction time through the muscle is shortened, the metabolic rate increased and the viscosity of the cell material diminished. Cooling the fluid reverses these changes.

THE ACTIVITIES OF THE MAMMALIAN HEART

Many of the experimental results just described for the frog's heart are true for the mammalian heart. Because of the thickness of the muscle perfusion of the chambers does not maintain the beat of the heart and it is necessary to perfuse the coronary vessels. If the perfusion is carried out with Locke's solution the excised rabbit heart can beat rhythmically for many hours.

Locke's Solution

NaCl	0·9 g.
KCl	0·042 g.
$CaCl_2$	0·024 g.
$NaHCO_3$	0·02 g.
Glucose	0·1 g.
Distilled water,	100 ml.
pH	7·4

This solution, warmed to 37° C and oxygenated, is perfused through a cannula tied in the aorta; the aortic valve closes and the fluid passes through the coronary vessels (Fig. 28, 8). If the fluid is warmed the heart beats more quickly and the conduction rate through the heart muscle and in the A.V. bundle is increased, but if the temperature is too high (above 44° C) the heart stops. Conversely if the perfusion fluid is cooled the heart rate falls.

The mammalian heart is much more sensitive to changes in pH of the perfusing fluid than is the frog's heart. Thus the dog's heart *in situ* stops suddenly in diastole about pH 6·0, apparently as the result of diminished contractility. If the pH of the fluid perfusing an isolated mammalian heart is reduced the rhythmicity of the S.A. node is diminished, conduction through the A.V. bundle is slowed and the ventricles and coronary arteries are dilated. Perfusion experiments also show that there is in the mammalian, as in the heart of the frog, a descending order of rhythmicity from the venous side to the ventricles.

Gross departures from the normal (3·5 to 5·5 m.equiv/l.) in the concentration of potassium in the blood plasma produce characteristic alterations in the electrocardiogram in man (Fig. 28, 55) and may result in sudden arrest of the heart. When the concentration of calcium is abnormally low the electrocardiogram also becomes abnormal (Q–T interval prolonged).

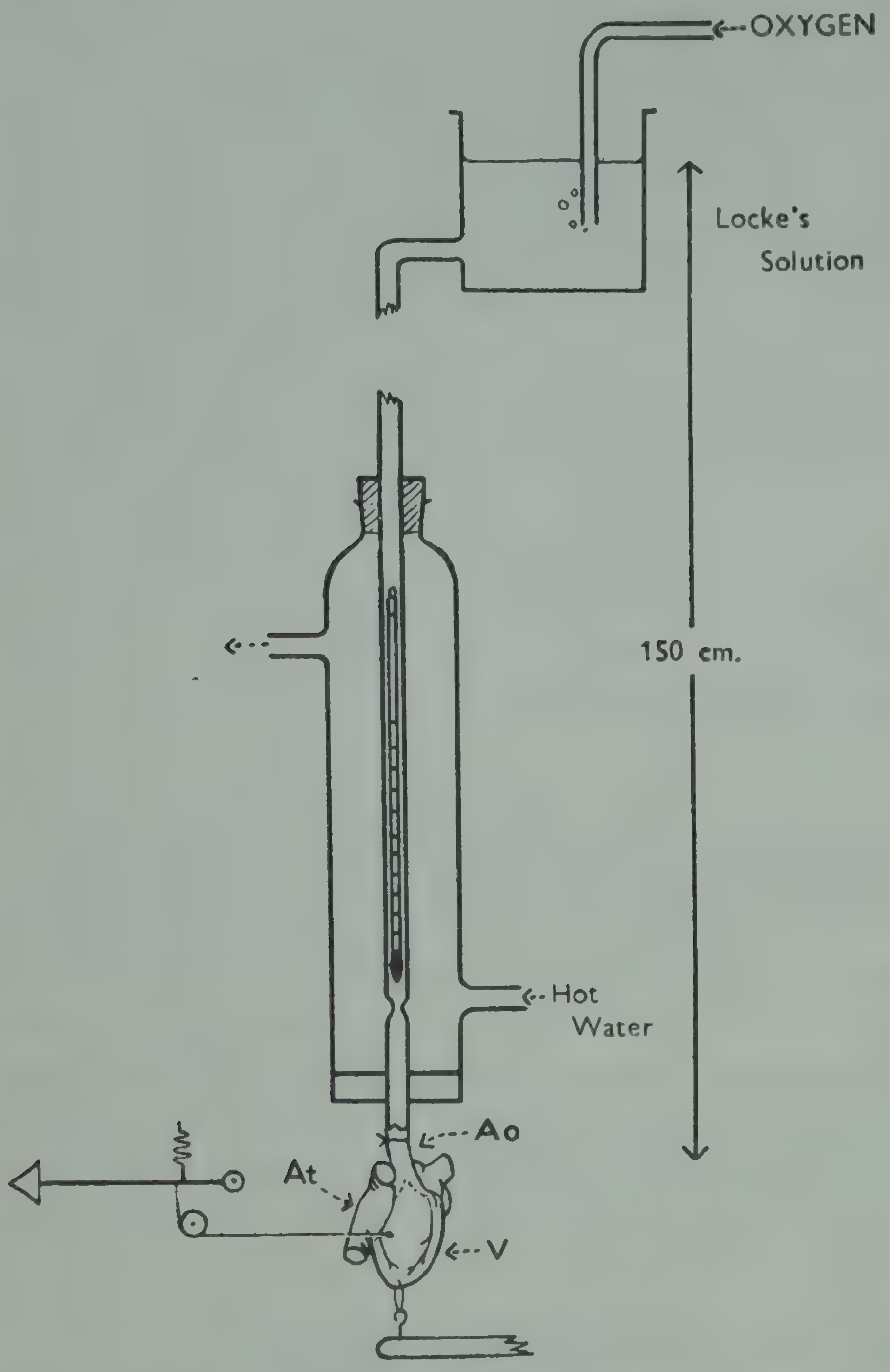


Fig. 28, 8

Diagram of the apparatus devised by Gunn for perfusion of the coronary circulation of the excised mammalian heart. The pressure of the fluid in the aorta closes the aortic valves so that the fluid enters the coronary arteries and drains from the coronary veins into the right side of the heart. Ao, aorta; V, ventricle; At, atrium.

Sequence of Events in the Cardiac Cycle

Since the heart rate in man is normally about 70 to 80 per minute, each cycle lasts about 0·8 sec. The atria first contract simultaneously (*atrial systole*) and expel their contents into the ventricles. Then after a slight pause the ventricles contract simultaneously (*ventricular systole*), expelling their blood into the pulmonary artery and the aorta. Ventricular systole is followed by a pause called *diastole*, during which all parts of the heart are relaxed. The time relations of the various events are shown in the diagrams (Figs. 28, 9 and 10), from which it may be calculated that the ventricles are contracting during nine hours out of the twenty-four, or sixty-three hours per week.

The sequence of events in the cardiac cycle is best studied from records of the pressure changes within the chambers of the heart. They take place, however, so rapidly that special manometers such as that devised by Wiggers (Fig. 28, 11) have to be used to record them accurately. This apparatus, mounted on a tube or trocar, is filled with anticoagulant fluid and the trocar is thrust through the heart wall of an animal into the chamber to be investigated. The changes in volume of the heart during the cardiac cycle may be recorded at the same time by means of a cardiometer (Fig. 28, 23). This is a glass vessel with two openings, one of which is connected to a volume recorder. The heart is inserted into the larger opening through an airtight rubber seal.

Let us consider first the curve of ventricular pressure (Fig. 28, 9). The contraction of the atrium drives blood into the relaxed ventricle through the open A.V. valves, causing a small rise of pressure in the ventricle. Just as the atrial contraction is dying away the ventricle suddenly contracts to produce a very rapid rise of pressure within this chamber. The flaps of the A.V. valves are pressed together by the rising intraventricular pressure and they are brought still closer together by the reduction in the diameter of the fibrous ring to which they are attached. At the same time the papillary muscles shorten, maintaining tension on the chordæ tendineæ so that the valve cusps cannot be everted into the atrium. For a very short time the ventricle is a closed cavity (isometric phase) but as the pressure rapidly rises above that in the aorta the semilunar valves open and there is a sudden increase of pressure in the aorta. Blood is ejected into the aorta at first quickly and then more slowly. The rate of outflow is shown by the steep fall of the curve of the volume of the ventricles. Approximately 0·3 sec from the beginning of systole the ventricle suddenly relaxes, there is a rapid fall in the pressure within it, and the aortic valve is rapidly closed by the higher pressure in the aorta.

The distension of the aorta by the ventricular ejection acting against the elastic wall causes this vessel to execute oscillations at its inherent frequency. The first of these oscillations, the *dicrotic wave,* is clearly visible on the falling curve of aortic pressure as the blood flows to the peripheral arteries. After the closure of the aortic valve the pressure in the ventricle falls rapidly (isometric phase) until it is less than the pressure in the atrium, when the A.V. valves open (end of isometric phase). The pressure in the ventricle continues to fall because in the intact animal, owing to the elastic traction of the lungs, the pressure in the thorax, and consequently in the ventricle during diastole, is less than atmospheric. In health the intraventricular pressure is never lower than the intrathoracic pressure and, therefore, the ventricle does not actually suck blood into itself. In the first third of ventricular diastole the inflow into the ventricle is very rapid, as indicated by the steep rise of the curve of ventricular volume, while during the second two-thirds the inflow is relatively slow.

The atrial pressure curve shows great variations in contour but there are always two main rises of pressure (Fig. 28, 9). Since there are no valves between the right atrium and the great veins these two alterations of pressure are transmitted to the great veins near the heart and may actually appear as visible pulsations in the external jugular vein in the neck—the so-called venous pulse (see Fig. 28, 14). The first rise in the atrial curve, *a*, is due to atrial systole; the second rise, *c*, is caused chiefly by

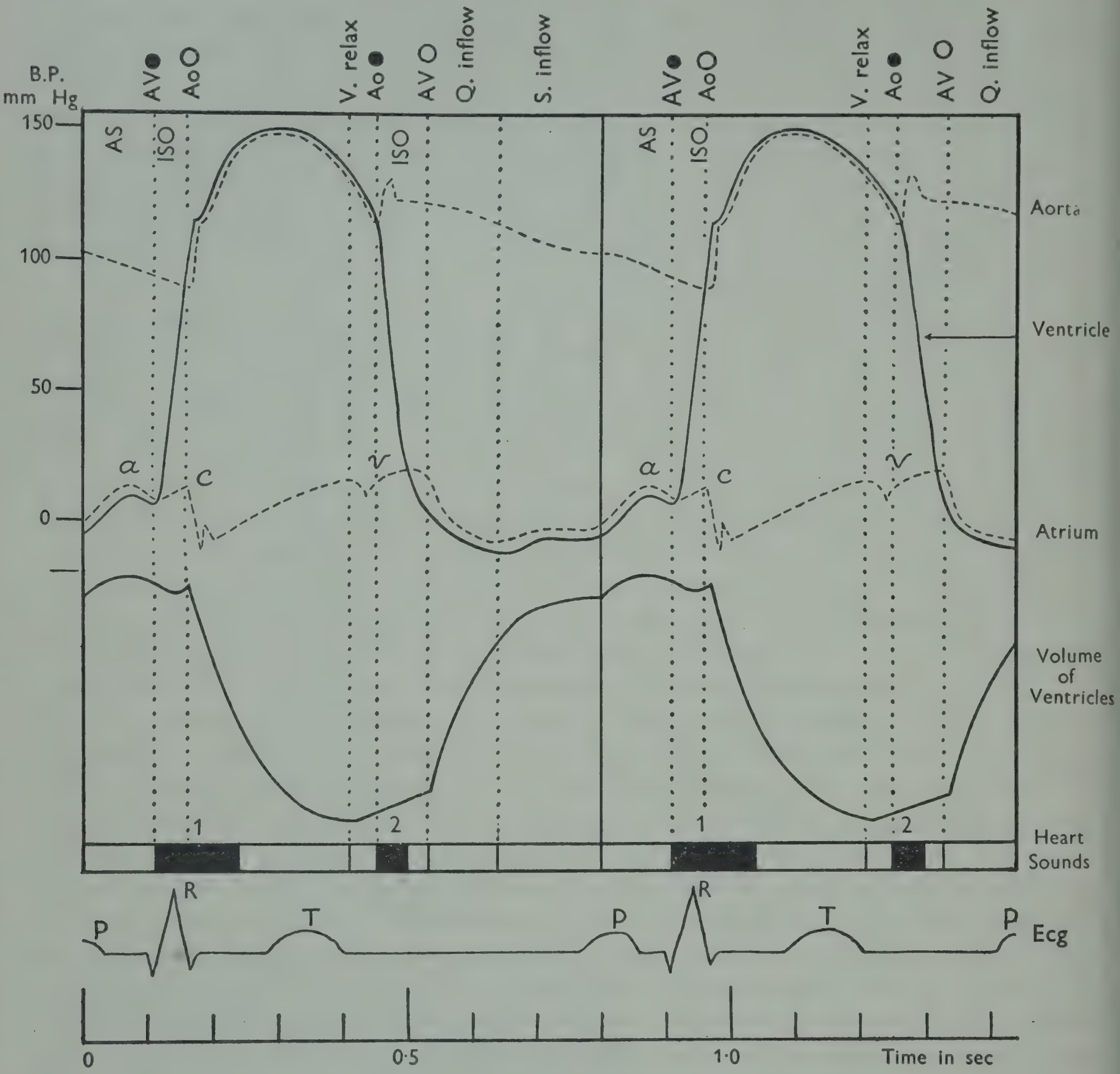


FIG. 28, 9

Curve of the pressure in the aorta, left ventricle, and left atrium constructed from data by Wiggers for a heart rate of 75 beats per minute. The volume curve of the ventricles, obtained by a cardiometer, rises when the volume increases. The timing of the occurrence of the first and second heart sounds is indicated by the black rectangles.

AS = atrial systole. V. relax. = ventricle relaxes.
ISO = isometric phase—both AV valves and aortic valves closed.
AV ● = AV valves closed. AV O = AV valves open.
Ao ● = aortic valves closed. Ao O = aortic valves open.
Q. inflow = quick inflow into ventricles. S. inflow = slow inflow into ventricles.

The changes in the right side are similar but the pressures are lower. (*After* C. J. Wiggers (1923), *Modern aspects of the circulation in Health and Disease*, 2nd ed. Philadelphia: Lea and Febiger.)

ventricular systole and closure of the A.V. valves below the atria and partly by the reduction in diameter of the A.V. fibrous ring. After this the atrial pressure usually falls, due to the pull on the A.V. rings by the contracting ventricle. Then the pressure rises again, *v*, as blood continues to flow into the atria while the A.V. valves are shut. The summit of the *v* wave is synchronous with the opening of the A.V. valves after which the pressure falls as blood flows into the ventricles.

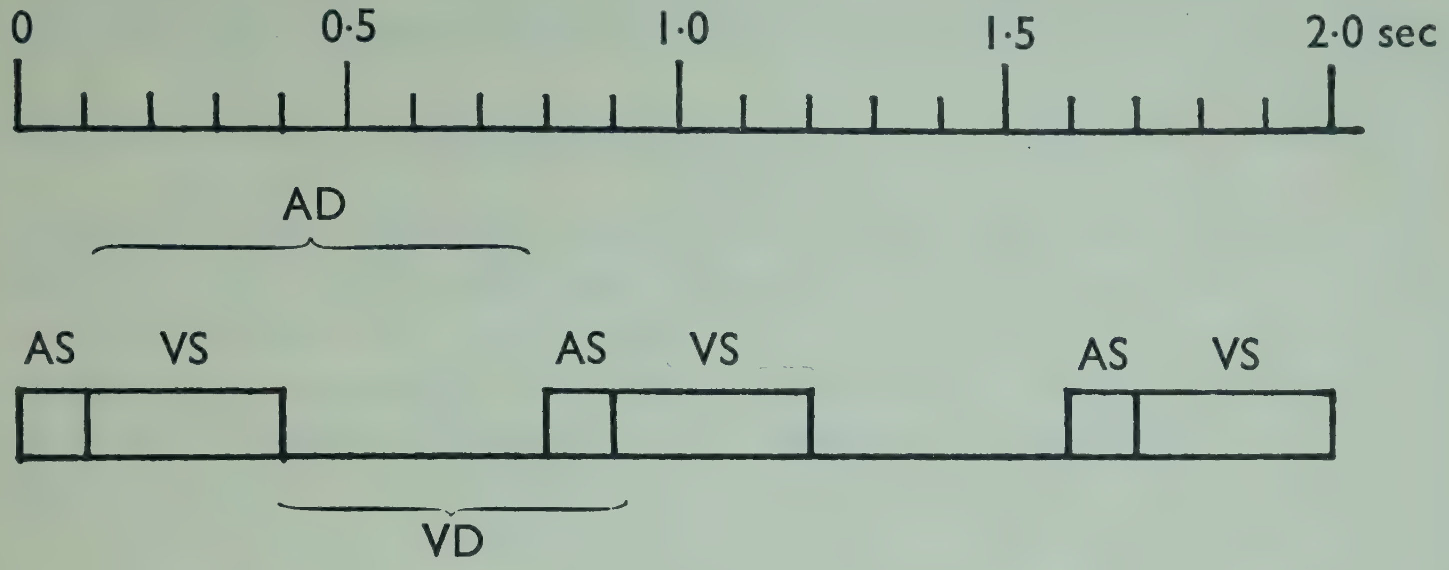

0
0·5
1·0
1·5
2·0 sec
AD
AS
VS
AS
VS
AS
VS
VD

FIG. 28, 10

Diagram drawn to scale to illustrate the events in the cardiac cycle when the heart is beating 75 times per minute.

AS = 0·1 sec (atrial systole).
AD = 0·7 sec (atrial diastole).
Atrial cycle = 0·8 sec
VS = 0·3 sec (ventricular systole).
VD = 0·5 sec (ventricular diastole).
Ventricular cycle = 0·8 sec

All parts of the heart are in diastole during the 0·4 sec between the end of VS and the beginning of AS.

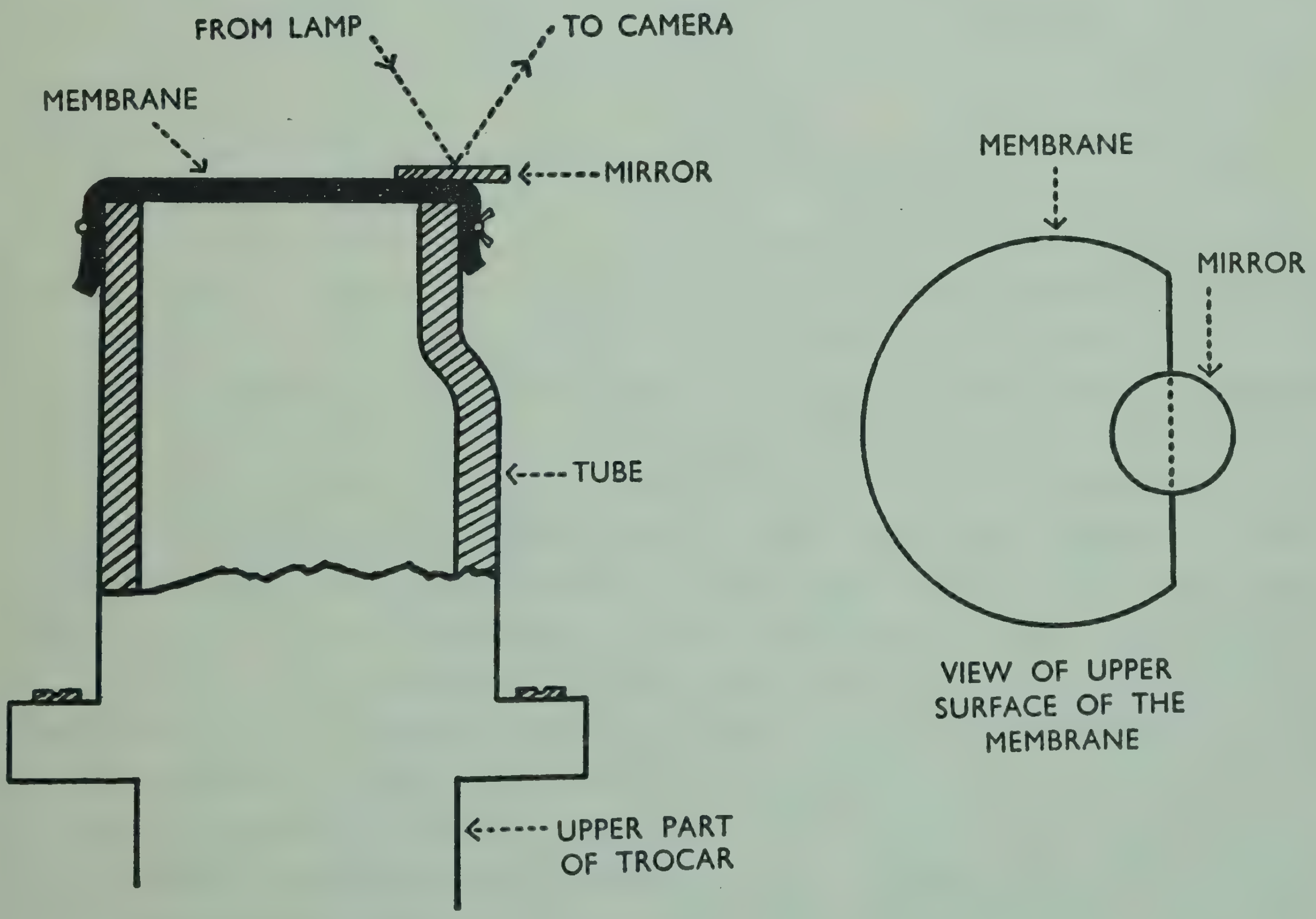

FROM LAMP
TO CAMERA
MEMBRANE
MIRROR
TUBE
UPPER PART OF TROCAR
MEMBRANE
MIRROR
VIEW OF UPPER SURFACE OF THE MEMBRANE

FIG. 28, 11

Wiggers' pattern manometer for recording rapid changes of pressure. The membrane is mounted on the end of a trocar, a sharp-pointed tube, which is thrust into one of the chambers of the heart. A rise of pressure within the tube bulges the membrane and tilts the mirror so that the beam of light is deflected (to the right in the diagram) across the slit of a recording camera.

The curves of pressure changes in the two sides of the heart are similar but those on the right are smaller in extent. Records of the pressure changes in the human heart, given in Figs. 28, 28 and 29 and also Fig. 29, 13, were obtained by attaching membrane manometers to catheters lying in the heart.

Return of Blood to the Heart

As we have seen, the heart, unlike most pumps, does not exert any suction on the entering fluid. Blood flows from the veins to the atria because the pressure in the veins is higher than that in the atria. The most important factor in returning blood to the heart is, of course, the action of the heart itself in forcing blood into a system of tubes with a limited capacity. If the capacity of the vascular system exceeds the ability of the heart to fill it as the result of excessive dilatation of the vessels of the splanchnic area, or of the skeletal muscles, the venous return is greatly diminished, the output of the heart falls and the circulation may fail.

The pressure within the chest, but outside the lung, that is the *intra-thoracic pressure,* is normally negative and this reduced pressure is exerted particularly on the thin-walled blood vessels, namely the veins and the atria. Accordingly blood flows from the abdomen and other parts where the pressure is above atmospheric into the thoracic veins where the pressure is below atmospheric. The effective filling pressure of the right side of the heart is actually the difference between the right atrial pressure and the intra-thoracic pressure, that is to say $0-(-5)=+5$ mm. Hg. The intra-thoracic pressure becomes more negative during inspiration and the filling pressure therefore higher. During inspiration the descent of the diaphragm causes an increase of intra-abdominal pressure which not only aids the return of blood to the thorax but also augments the flow of portal blood through the liver into the inferior vena cava.

The importance of this *respiratory pump* is shown in the performance of static effort, such as pulling on a tug-of-war rope. In this the glottis is closed and an expiratory effort is made so that the pressure in the thorax becomes positive (the Valsalva manœuvre). The return of blood to the heart is much reduced or ceases entirely, the superficial veins of the head and neck become engorged, the arterial blood pressure falls (Fig. 28, 40) so much that the diminished blood flow through the brain finally leads to giddiness and loss of consciousness. A similar state of affairs can occur during a severe spasm of coughing (*cough syncope*).

The return of blood to the heart is also assisted by the contraction of muscles all over the body, especially in the legs. When the muscles contract blood is squeezed out of the capillaries and smaller veins within the muscles into the larger veins between the muscles. This blood must flow towards the heart because the valves in the veins prevent any back flow. The alternate contractions and relaxations of the leg muscles which occur in walking serve to drive the blood back to the heart. A person standing rigidly at attention for a prolonged period may eventually faint because the impaired venous return to the heart by the *muscle pump* (Fig. 28, 12) results in a diminished output of the ventricles.

The size and distensibility of the pericardial sac obviously set a limit to the filling of the heart in diastole. If the sac is occupied with fluid or blood neither atria nor ventricles can dilate fully, and consequently the heart contains relatively little blood when it contracts. The output per beat is therefore diminished and a form of cardiac failure (cardiac tamponade) may follow. If the pericardial pressure rises above the venous pressure, as may happen after a stab wound of a ventricle, blood cannot enter the heart and the circulation ceases.

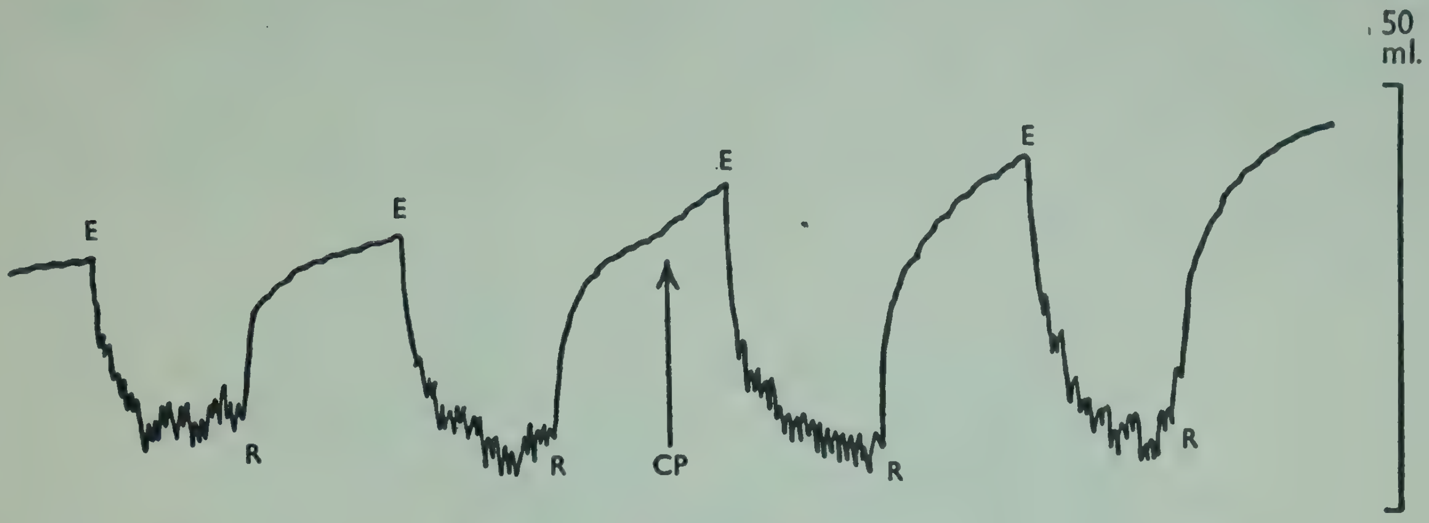


FIG. 28, 12

A record of the changes in volume of the calf of the leg measured with a plethysmograph (p. 502) to show the effects of the 'muscle pump'. A fall in volume is indicated by a downward movement of the trace. At E a pedal was pressed down once per second for 10 sec; R indicates the beginning of a rest period of 10 sec. At CP a cuff just above the knee was inflated to 90 mm Hg until the end of the trace. During each period of exercise blood was forced out of the calf and its volume diminished. Even when the venous pressure was raised to 90 mm Hg blood was forced out of the calf by the exercise. (H. Barcroft & A. C. Dornhorst (1949), *J. Physiol.* **109**, 402.)

MOVEMENTS OF THE HEART

The heart is quite soft in diastole, its shape being determined largely by the amount of blood within it and by the position of the body. When the ventricles contract the heart becomes hard and there is an increase in the antero-posterior, and a decrease in the transverse, diameter. This movement of the heart within the chest can often be felt as the *cardiac impulse* by the hand laid on the front of the chest.

If the thorax of an animal is opened and the heart exposed by incising the pericardium it can be seen that the apex rotates to the right at each systole. This movement is largely prevented in the intact animal by the attachment of the pericardium to the central tendon of the diaphragm. Leonardo da Vinci passed needles through the thoracic wall into the apical, middle and basal areas of the heart and found that at each ventricular systole the needle at the base moved most and the needle in the apex hardly moved at all. This very old experiment shows that at ventricular systole the base of the heart moves downwards and forwards while the apex is relatively fixed. The movement of the animal's atrio-ventricular ring towards the apex in systole has been confirmed by making X-ray photographs of a dog after marking the A.V. groove with silver wire. This movement is the result partly of the contraction of the ventricular fibres and partly of the elongation of the aorta and pulmonary artery which occurs at each systole. Because the apex of the heart lies against the chest wall in diastole, the hardening of the ventricles during systole, the tendency of the organ to rotate towards the right and the downward and forward movement of the base towards the apex all combine to produce, especially in young subjects, a palpable and sometimes visible pulsation in the fifth left intercostal space some three and a half inches from the mid-line of the sternum. In healthy persons this pulsation imparted to the chest wall during systole is no more than 1 to 1½ inches in extent. It is called the *apex beat*, although X-ray studies show that the actual apex lies about an inch lower down. Frequently the apex beat is not visible and often not palpable in the adult.

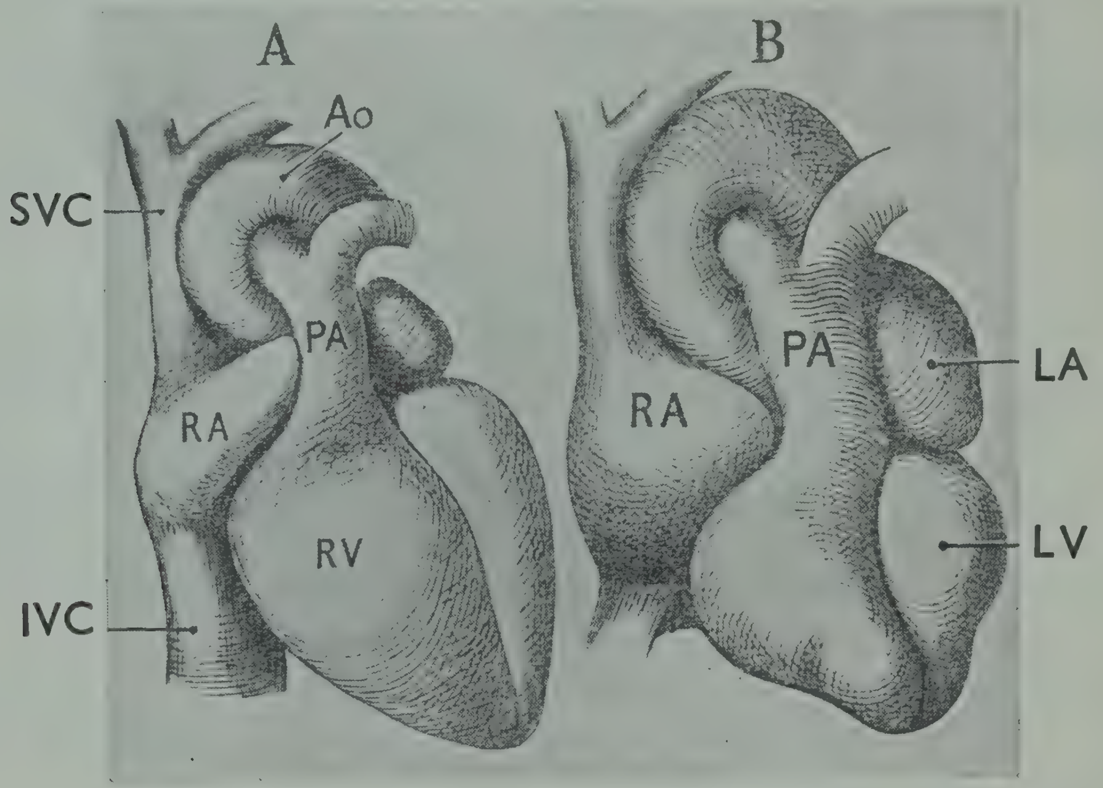


Fig. 28, 13

The appearance of the mammalian heart, A, at atrial systole and, B, at the end of ventricular systole. SVC and IVC, superior and inferior vena cava; Ao, aorta; PA, pulmonary artery; RA and LA, right and left atrium; RV and LV, right and left ventricle (Rollett (1880), *Hermann's Handbuch der Physiologie*, **4**, 170).

The movements of the heart just described and the ejection of blood into the aorta are accompanied by a recoil of the body, the vigour of which is related to the systolic discharge. Arrangements can be made to record the movements of the whole body reclining on a specially constructed table and the tracing so produced (*ballistocardiogram*) can then be analysed to give information about the stroke volume of the heart. The advantage of the ballistocardiograph is that it allows frequent measurements of cardiac output at rest to be made without any operative procedure. A disadvantage is that calibration of the graphic record is not easy.

THE HEART SOUNDS

Harvey originally observed that when the heart contracted it made a sound. Since his time clinicians have usually spoken of two heart sounds, the first at the beginning of ventricular systole and the second at the beginning of ventricular diastole. When the sounds are recorded by photographing the vibrations of a tambour placed on the thorax over the heart, or by means of a microphone connected to an oscillograph, four sets of vibrations can be detected. They are related in time to the venous pulse (see p. 497) as shown in the *phonocardiogram* illustrated in Fig. 28, 14.

The *first heart sound* is recorded as a prominent set of vibrations beginning just before the onset of the *c* wave, that is to say it is associated with ventricular systole. It is chiefly due to the sudden coming together of the cusps of the A.V. valves and the vibrations set up in them and in other structures by the sudden increase of pressure. The contraction of the muscle probably adds little to the first sound. The *second heart sound* is indicated by a set of vibrations at the beginning of ventricular diastole which coincides with a notch on the ascending limb of the *v* wave of the venous pulse trace (Fig. 28, 14). This sound is due to the closure of the semi-lunar cusps in the aorta and pulmonary artery; if one of the cusps is hooked back by a wire passed down a carotid artery into the aorta the second sound is replaced by a murmur due to the blood flowing back into the ventricle. A *third* set of vibrations of smaller amplitude corresponds to the last portion of the descending limb of the *v* wave of the jugular trace. Since it coincides with the end of the rapid filling phase of the ventricles it is believed to be due to disturbances set up by the sudden rush of blood into the relaxing ventricles. A *fourth* set of vibrations (A in

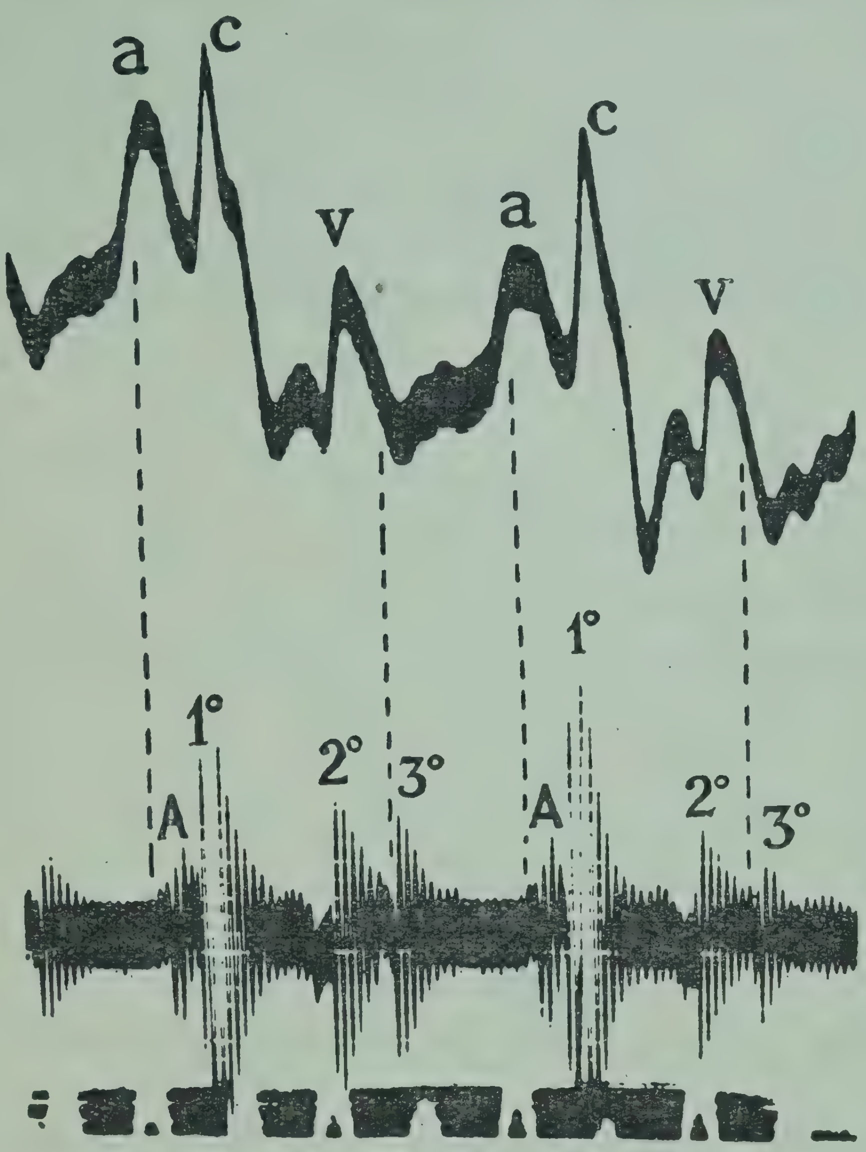


FIG. 28, 14

Top trace, jugular pulse. Middle trace, phonocardiogram. Lowest trace, time in 1/5 sec. The phonocardiogram is obtained by means of a Wiggers capsule placed on the chest wall over the heart. The atrial sound (A) appears 0·04 sec after the beginning of the *a* wave of the venous pulse. The third heart sound (3^0) coincides with the last portion of the descending limb of the *v* wave (end of the rapid-filling phase) (E. Braun-Menendez (1938), Heart Sounds, *Lancet*, ii, 762).

Fig. 28, 14) seen just after the beginning of the *a* wave of the venous pulse is due to atrial systole.

On listening over the heart, either by applying the ear directly to the chest wall or by means of a stethoscope (auscultation), at least two sounds are heard during each cardiac cycle. The first is low-pitched and of relatively long duration, while the second is higher in pitch and shorter. These two sounds (first and second sounds) correspond to the second and third sets of vibrations in the phonocardiogram. The rhythm (Fig. 28, 15) is best appreciated by listening to a slowly beating heart. When the heart is beating rapidly and diastole is therefore very short, it may be difficult to differentiate the two heart sounds, although the first sound can usually be distinguished because it coincides with the apex beat.

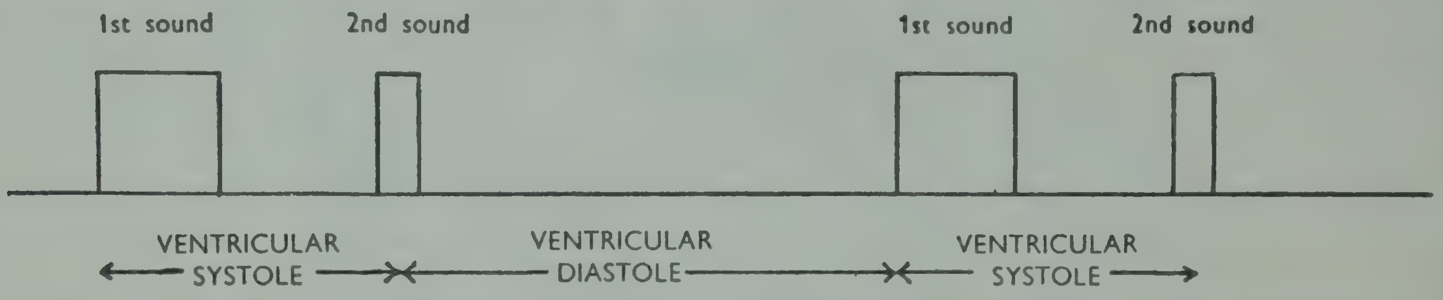


Diagram to illustrate the rhythm of the two cardiac sounds heard by auscultation.

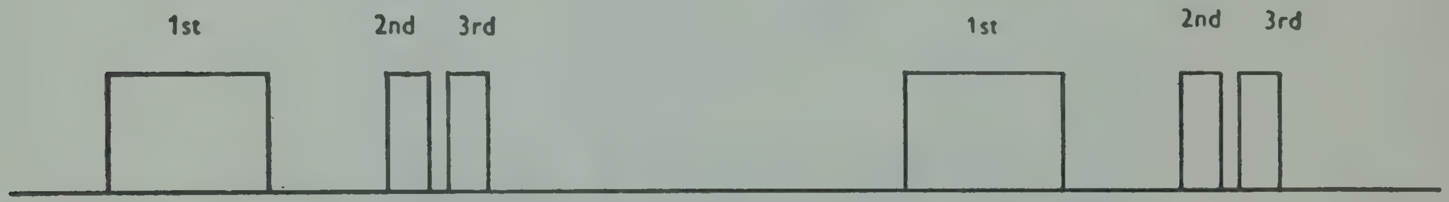


Fig. 28, 15

Diagram to illustrate the rhythm of the sounds when the third heart sound is audible.

The third heart sound is audible in 50 to 85 per cent. of healthy young persons. It occurs so closely after the second sound as to give rise to the impression that the second sound is reduplicated, the rhythm being as described in the lower part of Fig. 28, 15. The fourth sound is less frequently heard in health, appearing like a reduplicated first sound.

The flow of blood through the heart does not usually produce sounds other than those just described but, if the valves are leaky (incompetent) or narrowed (stenosed), abnormal sounds, called murmurs or bruits, may be heard or recorded in the phonocardiogram. No satisfactory explanation, however, has yet been given for the murmurs quite often heard during ventricular systole in apparently healthy people. Systolic murmurs are often heard in anæmia but, as they disappear when the anæmia is cured, they are probably connected in some way with the alteration in quality of the blood.

THE HEART-LUNG PREPARATION

The activity of the heart in the intact animal is continually being modified, not only by changes in pressure and resistance in the arteries and veins but also by impulses from the central nervous system. In order to study the behaviour of the isolated mammalian heart Starling, in 1910, devised the heart-lung preparation (Fig. 28, 16). The lungs are left in the circuit to act as oxygenators of the defibrinated blood which is pumped round the system.

An animal, usually a dog, is anæsthetized and a cannula is inserted into its trachea so that artificial respiration can be maintained by means of a pump. The chest is opened and a cannula is placed in one branch of the aorta, the other branches of which, with the descending aorta, are tied off. The superior vena cava is cannulated and connected to a reservoir containing defibrinated blood and the inferior vena cava is attached to a water manometer. In such a system blood can flow out of the left ventricle only by way of the coronary arteries and by the aortic cannula which leads to a mercury manometer recording mean arterial pressure. From here the blood is led into a bottle containing air which is compressed during ventricular systole, thus imitating the elasticity of the arterial walls and providing the pressure necessary to drive the blood on through the 'arterial resistance' during diastole. This artificial resistance consists of a thin-walled rubber tube contained within a glass tube, the space between the two tubes communicating with a reservoir of air under pressure. By varying the pressure in the reservoir the resistance to the flow of blood can be controlled. From here the blood passes through a warming coil before

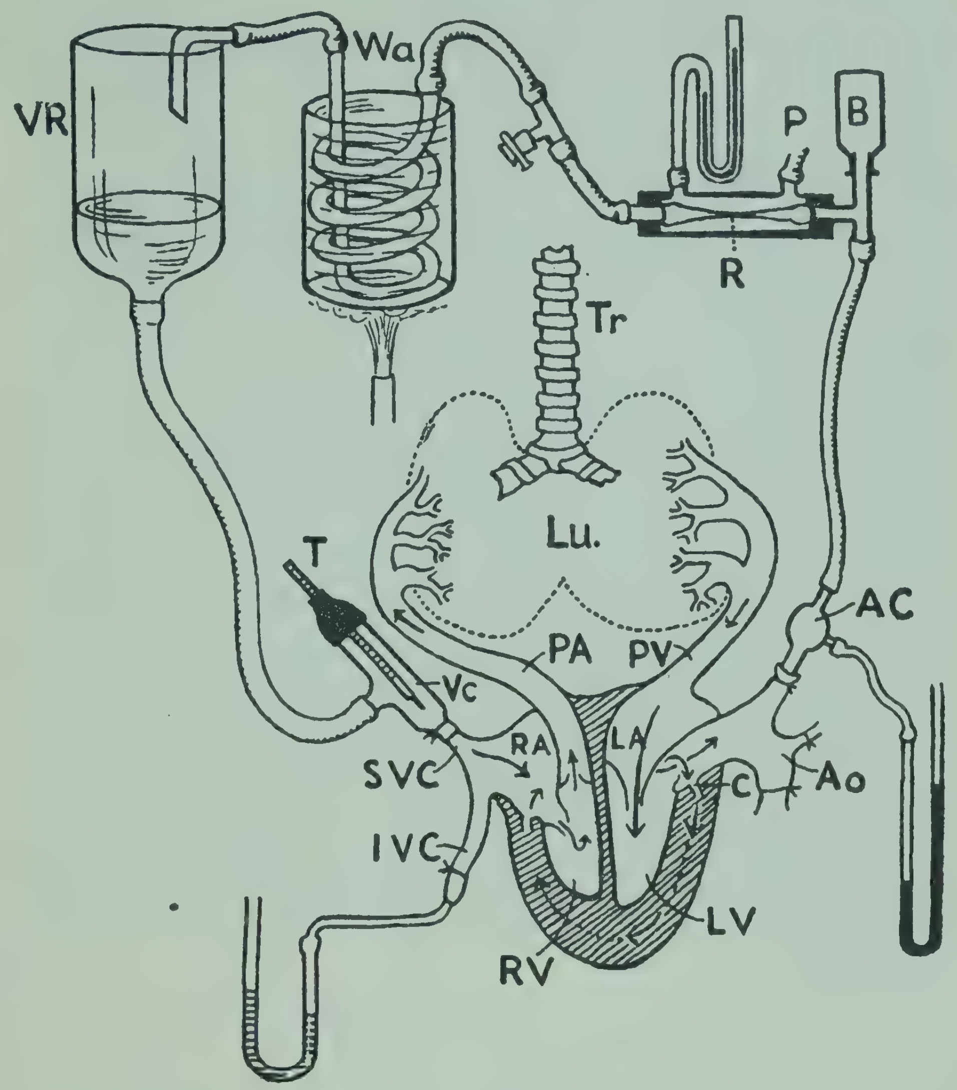


FIG. 28, 16

Diagram of circulation in heart-lung preparation. AC = arterial cannula. B = pressure bottle. R = arterial resistance. P = pump to control arterial resistance. Wa = warming coil. VR = venous reservoir. Vc = venous cannula. SVC and IVC = superior and inferior venæ cavæ. RA = right atrium. PA = pulmonary artery. Lu = lungs. PV = pulmonary vein. LA = left atrium. RV and LV = right and left ventricles. C = coronary blood supply. Ao = aorta. (*After* S. W. Patterson & E. H. Starling (1914), *J. Physiol.* **48**, 363, and Y. Bogue in C. Lovatt Evans' (1945), *Principles of Human Physiology*, 9th ed. p. 545. London: Churchill.)

being conducted to a venous reservoir from which it flows into the superior vena cava. The blood now passes into the right atrium and is driven by the right ventricle through the lungs where it is oxygenated, finally returning by the pulmonary veins to the left side of the heart.

The heart-lung preparation has yielded valuable information about the ability of the heart to act as a pump but the findings must be applied with caution to the heart of the intact animal. Further, unless the coronary sinus is cannulated no account is taken of the flow through the coronary vessels which is approximately 5 per cent. of the total output of the heart-lung preparation.

Venous Inflow.—With the heart-lung preparation it is found that if the venous inflow is kept constant the ventricular output is also constant, even if the arterial pressure increases from 60 mm Hg to 160 mm Hg or the heart rate increases from 80 to 160 beats per minute. If the venous inflow is altered there is an immediate and corresponding alteration in the ventricular output, as is shown by the experiment illustrated in Fig. 28, 17, where the arterial pressure was maintained constant. No change in heart rate occurred during this experiment since the cardiac nerves were cut and the central nervous system, not being perfused, was dead. At the first arrow the venous inflow was suddenly increased, the venous pressure rose and the ventricles became distended. At first the ventricles ejected less than they received so that their diastolic volume increased. This was followed by an increased output per beat until eventually the output equalled the input and no further increase in the heart volume occurred. During the part of the tracing marked B the ventricular volume at the end of systole was increased, that is, the ventricle no longer emptied so completely as during A. This new size was maintained until at the second arrow the inflow was suddenly reduced. The heart for a few beats expelled more blood than it received and the diastolic volume fell. Eventually a new equilibrium was reached when output again equalled inflow. In another experiment of this kind the output of a dog's heart weighing 50 g. was varied from 560 ml. to 3000 ml. per minute. It will thus be appreciated that the venous inflow is a most important factor in the regulation of the heart's activities.

Arterial Pressure.—The effect of varying the arterial pressure while the inflow to the heart-lung preparation is maintained constant is shown in Fig. 28, 18. At the first arrow the mean arterial pressure was suddenly increased from 98 to 128 mm Hg. During the first few beats after this change the ventricles did not discharge as much blood as they received and accordingly their diastolic volume increased. This was followed by a temporary increase in the output per beat, so that eventually the heart ejected all that it received and the diastolic and systolic heart volumes remained constant until the second arrow, when the mean arterial pressure was reduced. For a few beats the ventricles expelled more blood than they received and the diastolic volume was reduced.

Let us consider a heart-lung experiment in which the volume of the heart is being recorded with a membrane cardiometer. It may be found that the left ventricle is ejecting 10 ml. at each beat against an average arterial pressure of 100 mm Hg (range 90 mm Hg diastolic pressure to 110 mm Hg systolic pressure). At each ventricular systole when the ventricular pressure rises above 90 mm Hg the aortic valve opens and

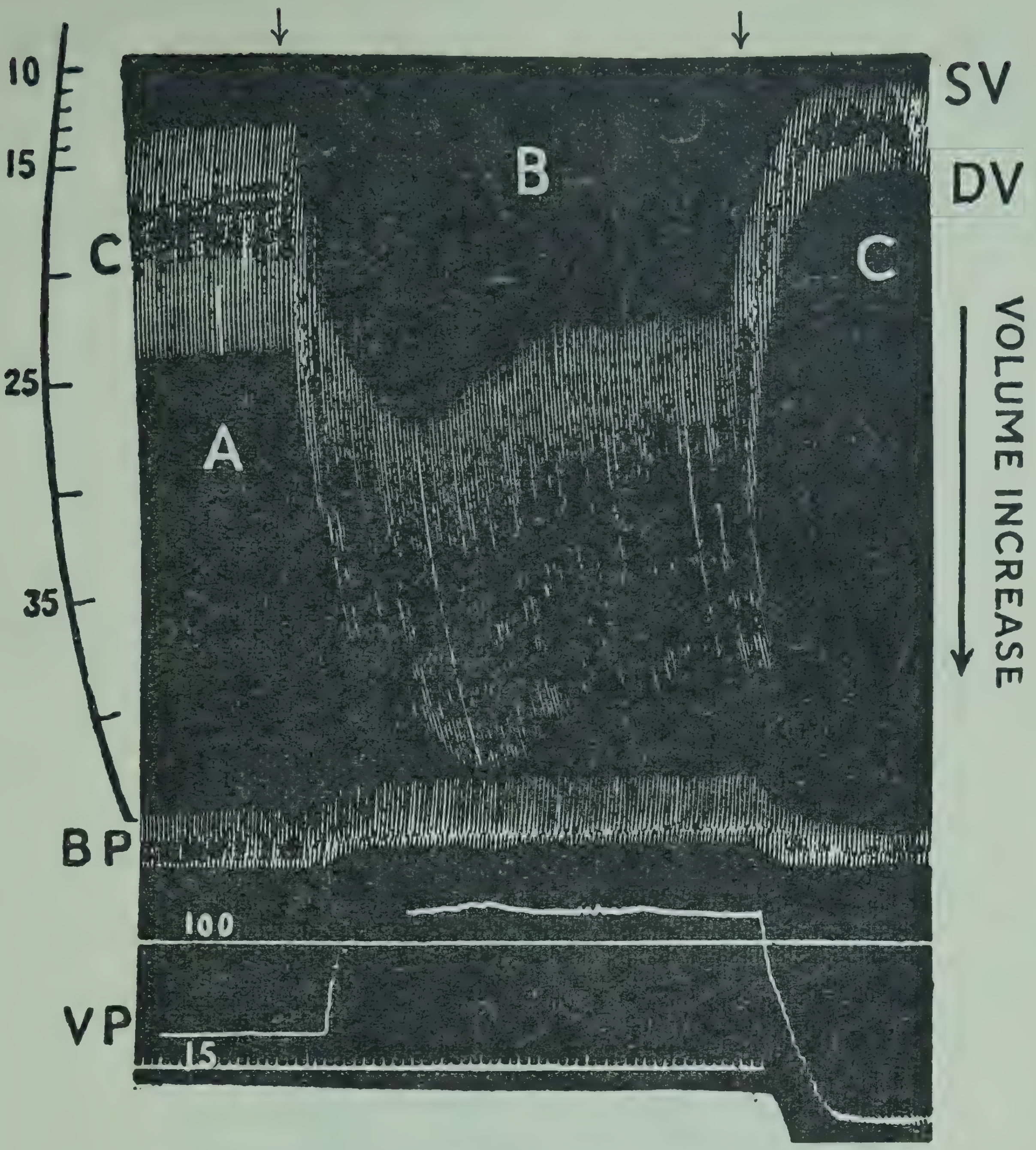


FIG. 28, 17

Effect of alteration in venous supply on volume of heart. Dog, 5·15 kg. Heart, 67 g. The venous pressure was varied by altering the height of the reservoir (VR in Fig. 28, 16) at the arrows.

Period	Venous inflow	Arterial pressure	Venous pressure	Output of heart	Rate per minute
	ml. per min	mm Hg	mm H_2O	ml. per min	
A	516	124	95	516	132
B	840	130	145	840	132
C	198	122	55	198	132

Time in sec

C = cardiometer curve with calibration curve in ml. on the left.
BP = arterial blood pressure.
VP = pressure in the inferior vena cava.
SV = systolic volume.
DV = diastolic volume.

Note.—Since an increase in volume is indicated by a downward line in the trace the volume of the ventricles at the end of systole is recorded by the upper limit of the trace and the volume at the end of diastole by the lower limit. The distance between these limits thus represents the systolic output per beat (stroke volume) (S. W. Patterson, H. Piper & E. H. Starling (1914), *J. Physiol.* **48**, 498).

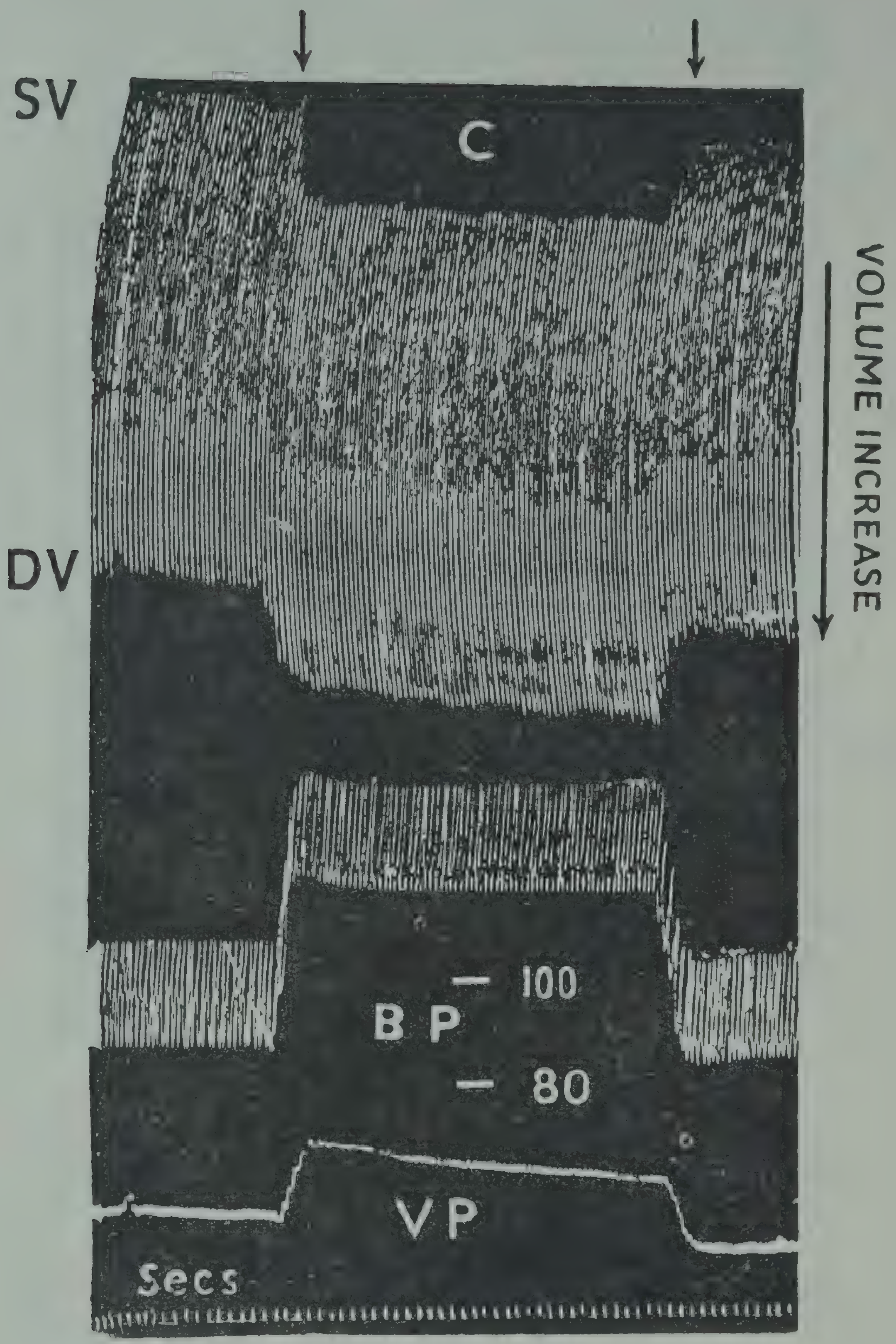


Fig. 28, 18

Tracing to show the effect of increased arterial pressure on the volume changes of the isolated heart with a steady inflow of 924 ml. of blood per minute.
Dog, 5·15 kg. Heart, 67 g. Output per beat 10 ml. throughout. The arterial pressure was altered by changing the peripheral resistance (R in Fig. 28, 16) at the arrows.

C = cardiometer curve.
BP = arterial blood pressure with scale in mm Hg.
VP = pressure in inferior vena cava.
SV = systolic volume.
DV = diastolic volume.

(S. W. Patterson, H. Piper & E. H. Starling (1914), *J. Physiol.* **48,** 483).

blood flows into the aorta, the aortic pressure rises quickly to its maximum value of 110 mm Hg and then falls slowly to 90 mm Hg. Suppose now that the arterial resistance is suddenly increased. The aortic pressure cannot fall so quickly as before, so that when the next ventricular beat occurs the aortic pressure is perhaps 100 mm Hg. As usual the semilunar cusps do not open till the pressure in the left ventricle is slightly higher than the pressure in the aorta (100 mm Hg in this case). Because of the higher diastolic pressure the aortic semilunar valves open slightly later and because of the higher aortic pressure and greater arterial resistance less blood is ejected by the left ventricle which therefore contains extra blood at the end of systole. In the meantime the venous inflow continues, so that during the following ventricular diastole the ventricle takes in its usual amount of blood in addition to the residuum from the previous contraction. It contains now, say, 12 ml. The next beat is observed to be more forcible and perhaps 9 ml. is ejected. In the course of three or four beats the original output of 10 ml. is regained and once more input is equal to output. During several beats, however, the heart has ejected less than 10 ml. per beat and it is therefore dilated, that is, its diastolic volume is slightly increased.

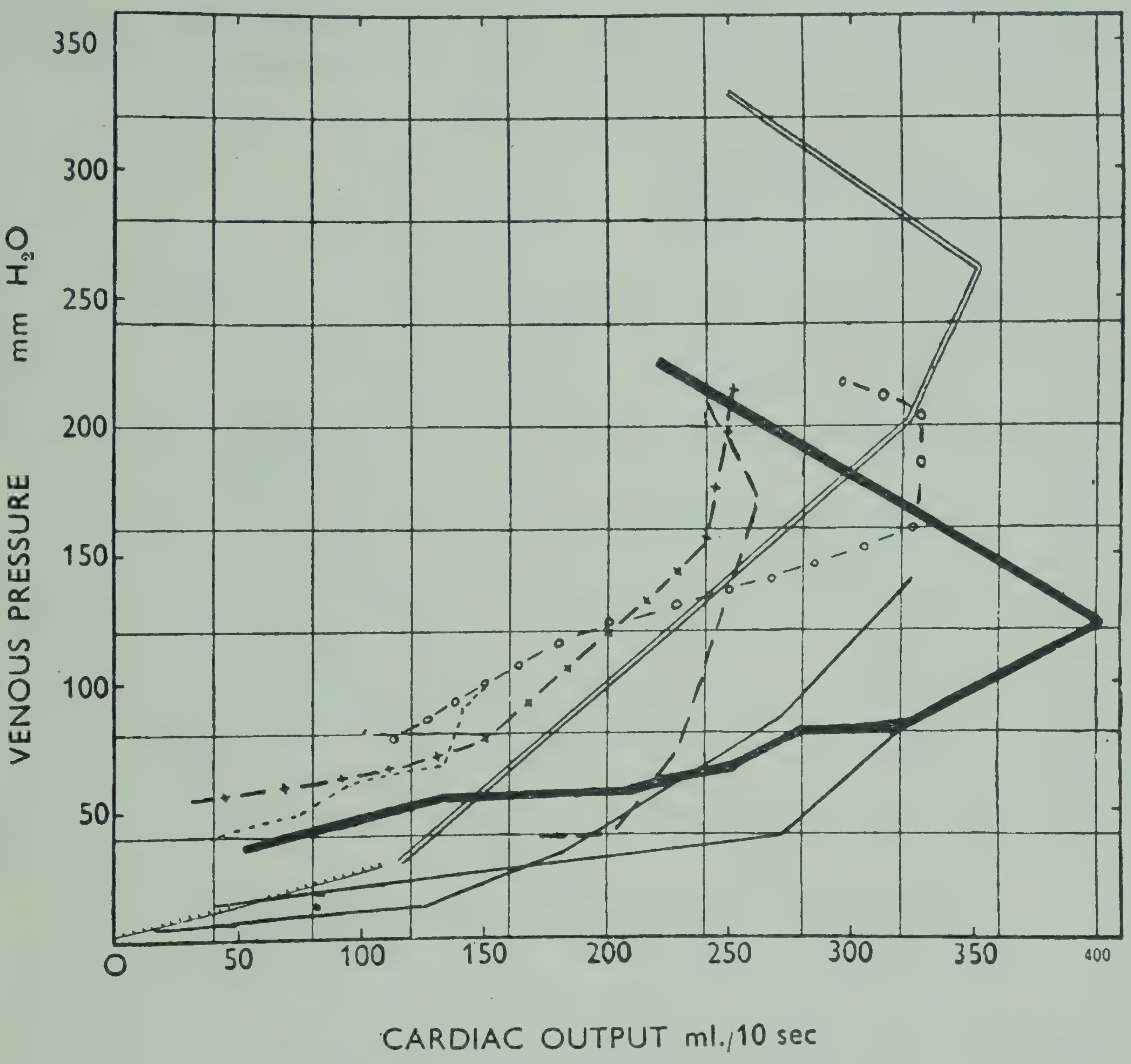


FIG. 28, 19

The relationship between the venous pressure on the right side of the heart and the cardiac output. Nine different experiments on the heart-lung preparation are plotted (S. W. Patterson, H. Piper & E. H. Starling (1914), *J. Physiol.* **48,** 367).

Starling's Law.—These results are described and summarized by Starling as the *law of the heart.* "The energy of the contraction is a function of the length of the muscle fibres." This is expressed graphically in Fig. 28, 19. It must not be imagined that the greater force of the contraction when the ventricle is distended is due to the greater tension in the muscle, because the diastolic pressure varies very little over a wide range of venous input, as is clearly shown by the first portion of the curves in Fig. 28, 19, where even a threefold increase (50 to 150 ml.) in output (and in inflow) caused practically no rise in venous pressure. Since an increase in the work of the isolated heart is accompanied by a dilatation there is a close correlation between diastolic size and oxygen usage. When the isolated heart fails it dilates, the atrial pressure rises, and there is a considerable increase in oxygen consumption.

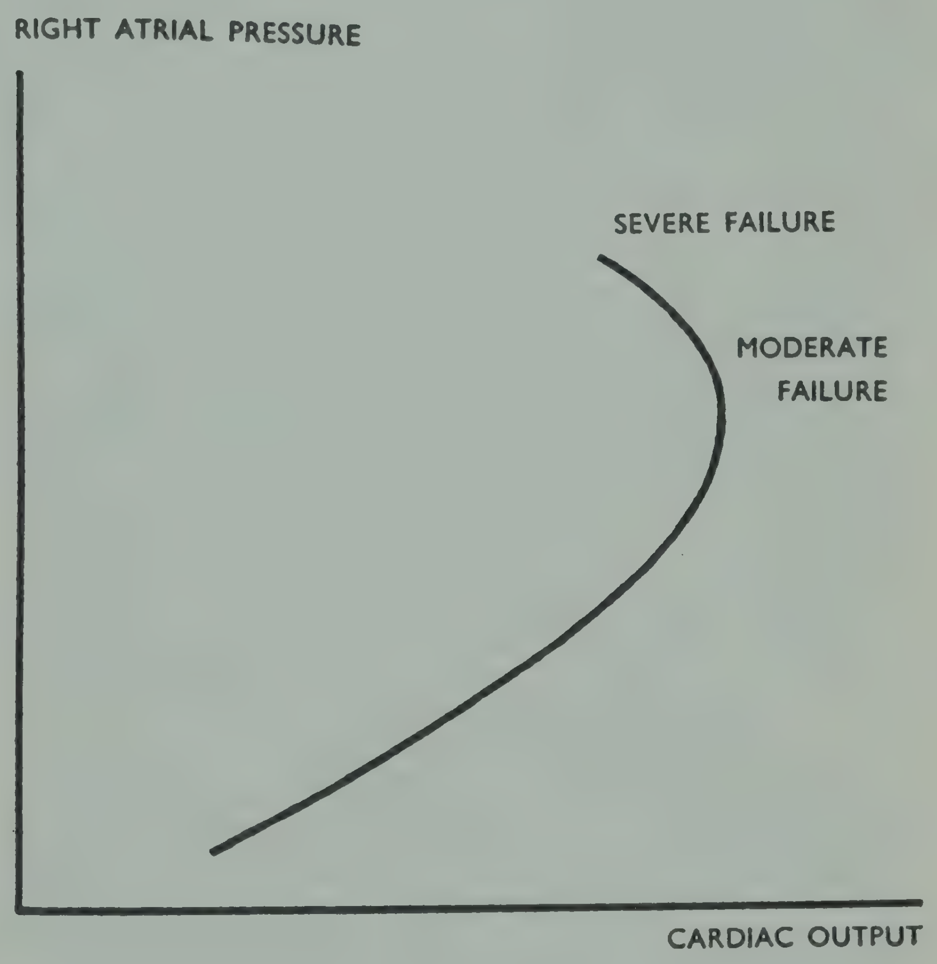


Fig. 28, 20

Diagram to illustrate how heart failure in man is explained on the basis of Starling's law. (*After* J. McMichael.)

Limits to the amount of dilatation that the heart can undergo are set by the pericardium and by the greater mechanical disadvantage at which the heart works when it is over-dilated. McMichael has drawn attention to the application of Starling's experiments to cardiac failure in man. As the filling pressure (right atrial pressure) rises the cardiac output increases to a point beyond which further dilatation is associated with a reduction in cardiac output (Fig. 28, 20). Cardiac failure is then present. In the early stages of cardiac failure the output at rest may not be much reduced but the ability of the heart to increase its output is diminished. In more severe failure the cardiac output at rest may be actually reduced.

THE WORK OF THE HEART

The work done by the left ventricle at each beat consists of two parts: (1) work done in forcing the blood out of the ventricle against the pressure in the aorta, (2) kinetic energy imparted to the blood.

If p is the mean pressure in the aorta and A is the cross-section of the aorta, then the force F acting across this area is $p\mathrm{A}$. If the stroke volume (output per beat) is V then the distance S through which the column advances is $\frac{\mathrm{V}}{\mathrm{A}}$. The work done is FS, the product of force and the distance through which it acts, that is $\mathrm{FS} = p\mathrm{A} \times \frac{\mathrm{V}}{\mathrm{A}} = p\mathrm{V}$. The kinetic energy is $\frac{1}{2}mv^2$ where m is the mass of the blood ejected at each beat and v is its velocity. The total work is, therefore, $p\mathrm{V} + \frac{1}{2}mv^2$. If all the measurements are made in c.g.s. units the work done can be expressed in ergs. The pressure p is usually given in mm Hg and $\frac{p \times 13\cdot6}{10}$ is therefore the pressure in cm of water. A column of water of this height and 1 cm² cross-section exerts a pressure of $\frac{p \times 13\cdot6 \times 981}{10}$ dynes/cm². For the purpose of this calculation the mean aortic pressure is not the average of the systolic and diastolic pressures but is the mean pressure acting during the time the aortic valves are open.

If ρ is the specific gravity of blood $m = \rho\mathrm{V}$; $\frac{1}{2}mv^2 = \frac{1}{2}\rho\mathrm{V}v^2$ ergs (if c.g.s. units are used). Now 1 kg. metre $= 1000 \times 100$ g. cm $= 10^5 \times 981$ ergs. Suppose that the heart output per beat is 60 ml., the mean arterial pressure is 100 mm Hg, the mean velocity is 50 cm/sec and $\rho = 1$. Then the work done by the left ventricle is

$$\begin{aligned} p\mathrm{V} + \tfrac{1}{2}mv^2 &= \frac{100 \times 13\cdot6 \times 981 \times 60}{10} + \tfrac{1}{2} \times 60 \times (50)^2 \text{ ergs} \\ &= \frac{100 \times 13\cdot6 \times 981 \times 60}{10 \times 10^5 \times 981} + \frac{\tfrac{1}{2} \times 60 \times (50)^2}{10^5 \times 981} \text{kg.m} \\ &= 0\cdot0816 + 0\cdot0007 = 0\cdot0823 \text{ kg.m} \end{aligned}$$

The work of the right side of the heart is much less, since the mean pressure in the pulmonary artery is only 15 mm Hg. The kinetic energy imparted to the blood is the same. The work of the right side is, therefore, 0·0122 + 0·0007 kg.m or 0·0129 kg.m. The total work done by the heart is then 0·0823 kg.m for the left ventricle and 0·0129 kg.m for the right ventricle, that is to say, approximately 0·1 kg.m per beat, or 10,000 kg.m per day for a man at rest. The kinetic factor becomes very much greater in exercise, when the heart output increases and the speed of ejection rises. In violent exercise the kinetic factor may amount to about a quarter of the total work of the heart, instead of only 1 or 2 per cent. in the above example, which is based on the data for a man at rest.

The analysis given above shows that the expression $p\mathrm{V}$, that is the product of the mean pressure and the output volume, forms a good approximation to the work of the heart. For clinical puposes, therefore, it is justifiable to use the formula,

Cardiac work = Cardiac output × Mean arterial pressure which emphasizes the important fact that the work of the heart is increased by factors elevating either its output or the pressure against which it has to operate.

Efficiency of the Heart.—In a heart-lung preparation it is easy to vary the work done and to calculate it fairly exactly by the method just described. At the same time it is possible to measure the oxygen consumption of the heart either (1) by analysing the expired air and making a small allowance for the oxygen consumption of the lungs, or (2) by measuring the difference in oxygen content between the aortic blood and the coronary outflow and making due allowance for the volume of the blood returning by other veins (p. 426).

Table 28, 21

Heart-lung Preparation

Denervated (Evans and Matsuoka, 1915)

	Aortic Pressure mm. Hg.	Output ml./min.	Rate min.	Work kg. m./min.	O_2 used ml./min.
Pressure raised	40	282	130	0·18	1·63
	80	280	130	0·34	1·98
Output raised	80	282	130	0·35	2·33
	80	562	130	0·78	2·64

Innervated Heart (Gollwitzer-Meier *et al.*, 1938)

Pressure raised	71	550	125	0·62	3·5
	136	576	97	1·25	3·2
Output raised	60	630	154	0·60	3·8
	63	1128	156	1·28	5·4

From C. Lovatt Evans (1939), *Edinb. med. J.* **46**, 733.

Knowing that 1 litre of oxygen corresponds to about 2000 kg.m of energy liberated we can find the ratio between work done and energy used—that is, the efficiency of the heart as a machine. Cardiac catheterization has recently made it possible to estimate the efficiency of the human left ventricle and the average value has been found to be 23 per cent.

If the work of the heart in the heart-lung preparation is increased by raising the aortic pressure, or by an increase in the output, the rate at which the heart uses oxygen also rises. The efficiency, however, is found to increase gradually from some 5 to 10 per cent. up to 20 to 25 per cent. When the work done becomes excessive the oxygen consumption increases much more rapidly and the efficiency declines.

The heart *in situ* with its nerve supply intact seems to work more efficiently than the isolated heart because an increase in the work done by the heart does not necessarily raise its oxygen consumption. This is illustrated in Table 28, 21, which summarizes experiments in which the work of the heart was approximately doubled.

If adrenaline is added to the blood flowing through the isolated heart the rate and strength of the beat are increased, but at the same time the oxygen consumption rises even more, so that the efficiency of the heart declines. Slow continuous intravenous infusion in man (0·1 to 0·4 μg. per kg. per min) produces a greatly augmented cardiac output and increased oxygen consumption. Infusion of noradrenaline at similar rates, however, does not have these effects.

SYSTOLIC OUTPUT OF THE HEART

An important but technically difficult problem is the determination of the amount of blood ejected from the ventricles at each beat of the heart. The output per beat of each ventricle is called the *stroke volume*. We have already seen that the stroke volumes of the two ventricles cannot be different for more than a few moments or blood would accumulate somewhere in the vascular system.

In animals it is possible to enclose the heart in a rigid box, or cardiometer (Fig. 28, 22) connected by tubing to a volume recorder. If the apparatus responds quickly and correctly the difference between maximum volume at diastole and minimum volume in systole gives the combined output of the two ventricles; that is, twice the stroke volume. The disadvantage of this method is that the chest has to be opened to get the cardiometer into position.

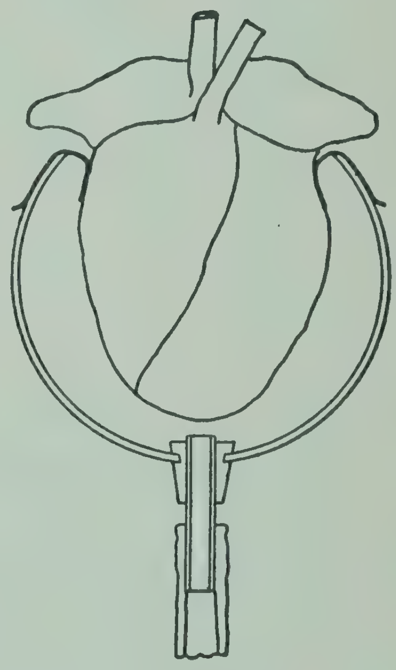


FIG. 28, 22

The cardiometer for measuring changes in volume of the exposed heart (Yandell Henderson (1906), *Amer. J. Physiol.* **16**, 325.

Fick Principle.—For the determination of the cardiac output in man only indirect methods are applicable. Most of these depend upon the principle enunciated about sixty years ago by Fick who pointed out that in a given time the total amount of any gas gained or lost in the lungs must be equal to the difference between the amounts of the gas brought to the lungs in the arterial blood and the amount leaving the lungs in the venous blood. Thus if we can measure (*a*) the oxygen content of the arterial blood and (*b*) the oxygen content of the blood in the pulmonary artery (mixed venous blood) and (*c*) the amount of oxygen passing from the lungs into the blood per unit of time, then the output of the heart is easily calculated. An average value for (*a*) is 19 ml. per 100 ml. blood, for (*b*) is 14 ml. per 100 ml. blood, for (*c*) is 250 ml. per minute. Each 100 ml. of blood in passing through the lungs gains 5 ml. of oxygen, that is, the arteriovenous difference. Since 250 ml. of oxygen actually passes into the blood (see Fig. 28, 23)

$$\frac{250}{5} \times 100 \text{ ml.} = 5000 \text{ ml. or 5 litres}$$

of blood are needed to carry that amount of oxygen, and this is the cardiac output per minute. If the heart contracts 72 times per minute then the stroke volume of each ventricle is 5000 ÷ 72 = 70 ml. Although this method is simple in principle it is only quite recently that it has been found possible to use it clinically. The value of (*a*) can be got by analysis of arterial blood but the value of (*b*) cannot be obtained by analysis of blood from any easily accessible vein, say a vein in the arm, because its oxygen content differs from that of the blood in the pulmonary artery which contains mixed venous blood from the whole body.

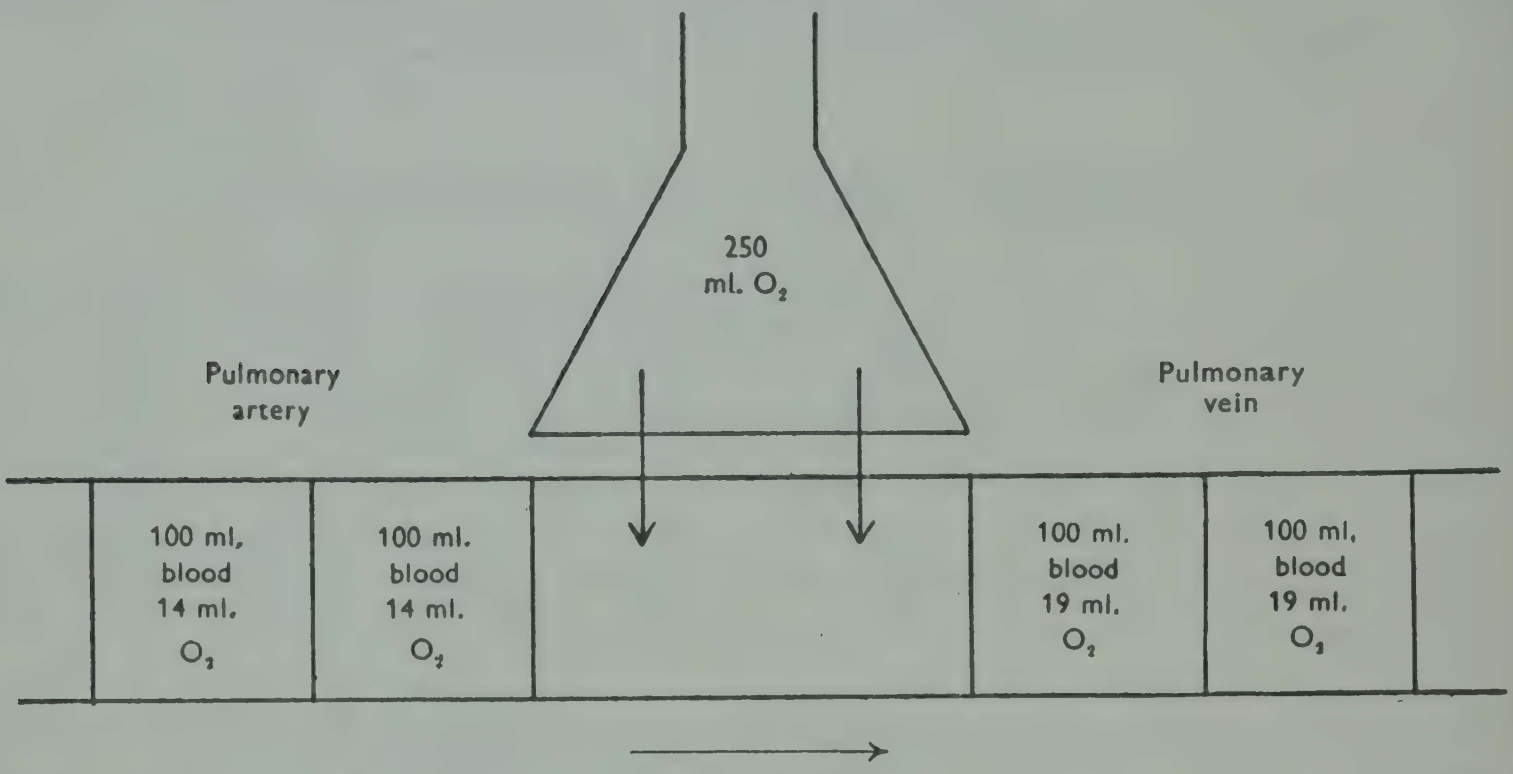


FIG. 28, 23

Scheme to illustrate the determination of cardiac output by the Fick principle. Each rectangle represents 100 ml. of blood which gains 5 ml. of oxygen in passing through the lungs. 250 ml. of oxygen disappears from the lungs per minute. This volume would be accounted for by $\frac{250}{5}$ or 50 rectangles. The flow is therefore $\frac{250}{5} \times 100$ ml., or 5 litres per minute.

Cardiac Output in Man.—Measurement of cardiac output in man can now be made by the technique described by Cournand in 1941. A fine semi-rigid catheter is introduced into a vein at the bend of the elbow and pushed on until its tip is in the right ventricle or preferably in the pulmonary artery. Through this catheter a 15 ml. sample of mixed venous blood is withdrawn under oil for estimation of oxygen content. A sample of arterial blood may be obtained by puncture of the femoral artery with a fine needle attached to a syringe. This procedure can be avoided, when the lungs are normal, by estimating the total oxygen capacity of the venous sample and then assuming that the arterial blood is 95 per cent. saturated (see p. 576). The oxygen consumption is determined at the same time by a spirometer or by the Douglas bag technique (see p. 162).

The results obtained in this way represent the effective output of each ventricle but do not include any run-back of blood into the ventricles such as may occur in valvular disease. If all the oxygen taken up from the inspired air is carried in the blood and is not utilized in the lungs the results should be correct for normal persons. The average arteriovenous oxygen difference at rest is 4·5 ml. O_2 per 100 ml. blood. Since the average oxygen consumption is 240 ml. per minute the average cardiac output is $\frac{240}{4\cdot5} \times 100 = 5\cdot3$ litres per minute. The range in normal adults under basal conditions is from 4 to 7 litres per minute, with an average of 5·3 litres. This corresponds to stroke volumes between 70 and 80 ml. The cardiac output falls by about one-third when the subject is moved from the recumbent to the erect position. In severe anæmia it has been found that the arteriovenous oxygen difference may be only 2 ml. per 100 ml. of blood. If the oxygen usage is 240 ml. per minute then the cardiac output is $\frac{240}{2} \times 100 = 12$ litres per minute. This is a considerable increase in cardiac output which is invariably associated with a high pressure in the right atrium. The cardiac output in exercise may be increased to as much as 30 litres per minute, partly by an augmented stroke volume and partly by an increase in cardiac rate. Apprehension and anxiety may be accompanied by an increase of some 10 to 20 per cent. over the basal value. The output is some 10 per cent. lower during sleep.

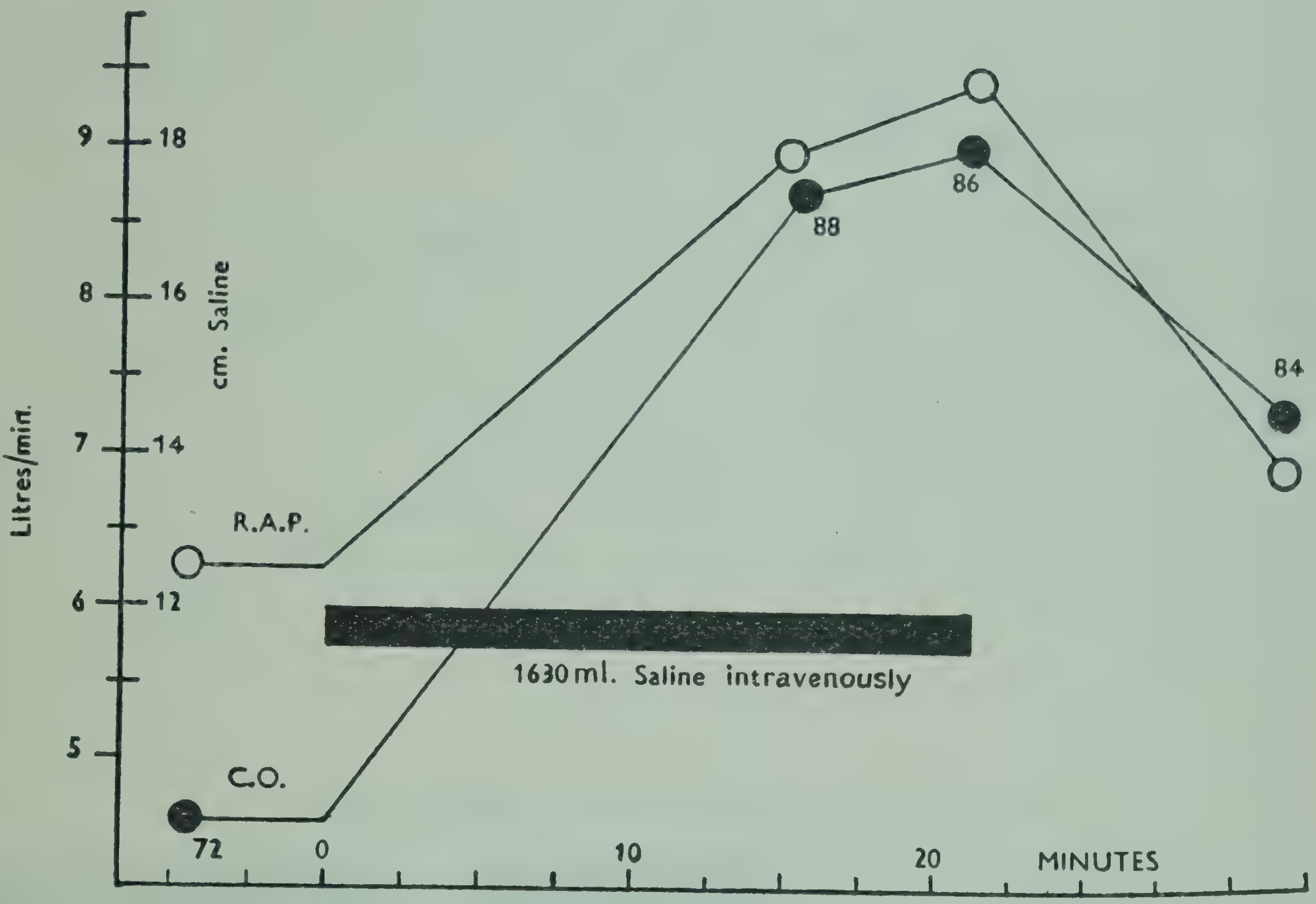


Fig. 28, 24

Effect of raising right atrial pressure (R.A.P.) by saline infusion. During infusion the cardiac output (C.O.) rises. On stopping the infusion saline is lost from the circulation and right atrial pressure and cardiac output fall together. The figures beside the cardiac output determinations indicate heart rate (J. McMichael & E. P. Sharpey-Schafer (1944), *Brit. Heart J.* **6**, 35).

The normal cardiac output is closely related to body surface area from which it can be predicted with an accuracy of about 10 per cent. For this reason cardiac output is often expressed as the *cardiac index*, that is, the minute volume per sq. metre of body surface. The normal cardiac index is 3·1 to 3·2 litres per sq. metre.

By the cardiac catheterization technique Starling's law has received experimental confirmation in man. When the right atrial pressure is increased by infusion of saline the cardiac output increases (Fig. 28, 24) and conversely when the right atrial pressure is lowered by the withdrawal of blood the cardiac output falls (Fig. 28, 25). The nature of the relationship is illustrated in Fig. 28, 20, from which it will be seen that the cardiac output increases with the right atrial pressure up to a certain value, beyond which any further increase of atrial pressure leads to a reduction in output.

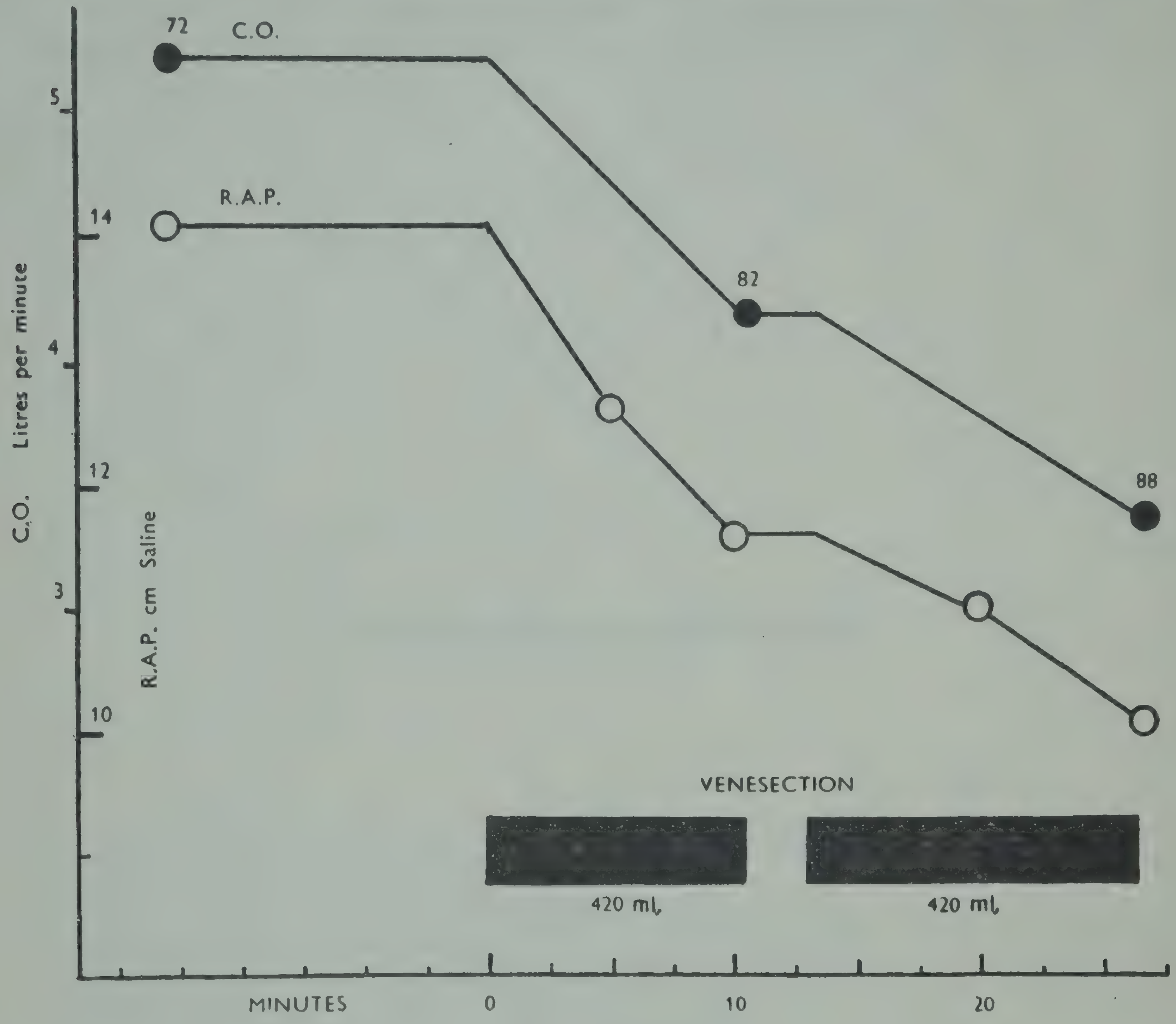


Fig. 28, 25

Effect of venesection on cardiac output and right atrial pressure. The figures beside cardiac output determinations indicate heart rate. Right atrial pressure (R.A.P.) and cardiac output (C.O.) fall in parallel (J. McMichael & E. P. Sharpey-Schafer (1944), *Brit. Heart J.* **6**, 34).

When a rise in the right atrial pressure produces little or no further increase in cardiac output cardiac failure is present; in extreme cardiac failure a further increase in right atrial pressure leads actually to a fall of cardiac output. The distension of the veins and the increased venous pressure in cardiac failure is not necessarily a purely passive phenomenon due to damming back of blood which the heart cannot deal with; it may rather represent a compensatory mechanism, the disabled heart requiring a greater head of pressure to perform its work.

A further advantage of cardiac catheterization is that it allows measurements of intracardiac and pulmonary artery pressures to be made in man and recorded by the optical methods previously applicable only to animals. Fig. 28, 28, and Fig. 28, 29, are examples of such records. They should be compared with Fig. 28, 9, which is based on animal experiments.

Now that accurate figures are available it is possible to see how the cardiac output is distributed to the various organs and tissues in man under resting conditions:—

	Litres per min
Digestive viscera (flow in hepatic vein)	1·5
Kidneys	1·2
Brain	0·9
Extremities	1·8
	5·4

The proportion of the cardiac output normally passing to the skin and endocrine organs is therefore very small; that passing through the kidneys is very large.

The cardiac output in man may be calculated from the ballistocardiogram (p. 438). It may also be measured by injecting a known amount (X mg.) of a non-diffusible dye into a vein and following the rapid rise and fall of its concentration in arterial blood obtained by needle puncture of, say, the brachial artery. The falling concentration, before recirculation of the dye begins, can be plotted as a graph which, if extrapolated to zero concentration, will give the time after injection at which all the dye must have been ejected from the heart if there had been no recirculation (Y seconds). If the average concentration of the dye in the blood is calculated for this period (Z mg./litre) then $\left(\frac{X}{Z}\times 100\right)$ litres is the amount of blood required for the ejection of X mg. of dye. Since this took place in Y seconds the cardiac output in litres per min must have been $\left(\frac{X}{Z}\times 100\right)\times\frac{60}{Y}$.

CONDUCTION THROUGH THE HEART

We have seen that in the frog an excitatory wave can be transmitted all over the heart from any point of stimulation. In the mammalian heart, however, the muscular continuity between atria and ventricles is interrupted at the fibrous atrioventricular rings and impulses are transmitted from atria to ventricles through a specialized conducting tissue. Kent and His discovered that there is a distinct bundle 1 to 2 mm in cross section, called the atrioventricular or A.V. bundle, of primitive muscle fibres which begins at the atrioventricular node at the base of the atrial septum and passes along the top of the ventricular septum (Fig. 28, 26). The A.V. bundle then divides into two branches (fasciculi) which run down, one on each side of the ventricular septum, to the right and left ventricles, each branch giving off very numerous twigs to form a reticulated layer of tissue just under the endocardium of both ventricles, called Purkinje tissue from which small branches penetrate the muscle to reach the pericardium. The cells of the Purkinje tissue show striation only in the periphery and have a high glycogen content. The fibres of the A.V. bundle and its branches cannot be distinguished from ordinary heart muscle fibres on dissection but can be traced microscopically. In some animals, but not in dog or man, there is a distinct fibrous sheath round the A.V. bundle which makes it possible to delineate the bundle and its branches by injecting coloured materials into this sheath. A large number of nerve fibres (but no ganglion cells in man) run alongside the A.V. bundle but they do not seem to be concerned with the propagation of the excitatory impulse. If the A.V. bundle is cut the atria, excited from the sinus, continue to beat at their usual rate but the ventricles beat at a much slower rate and quite independently of the atria, the site of impulse formation being an idioventricular centre distal to the block in the A.V. bundle. This phenomenon, called *complete heart block*, may occur in man when conduction through the bundle is completely interrupted by disease.

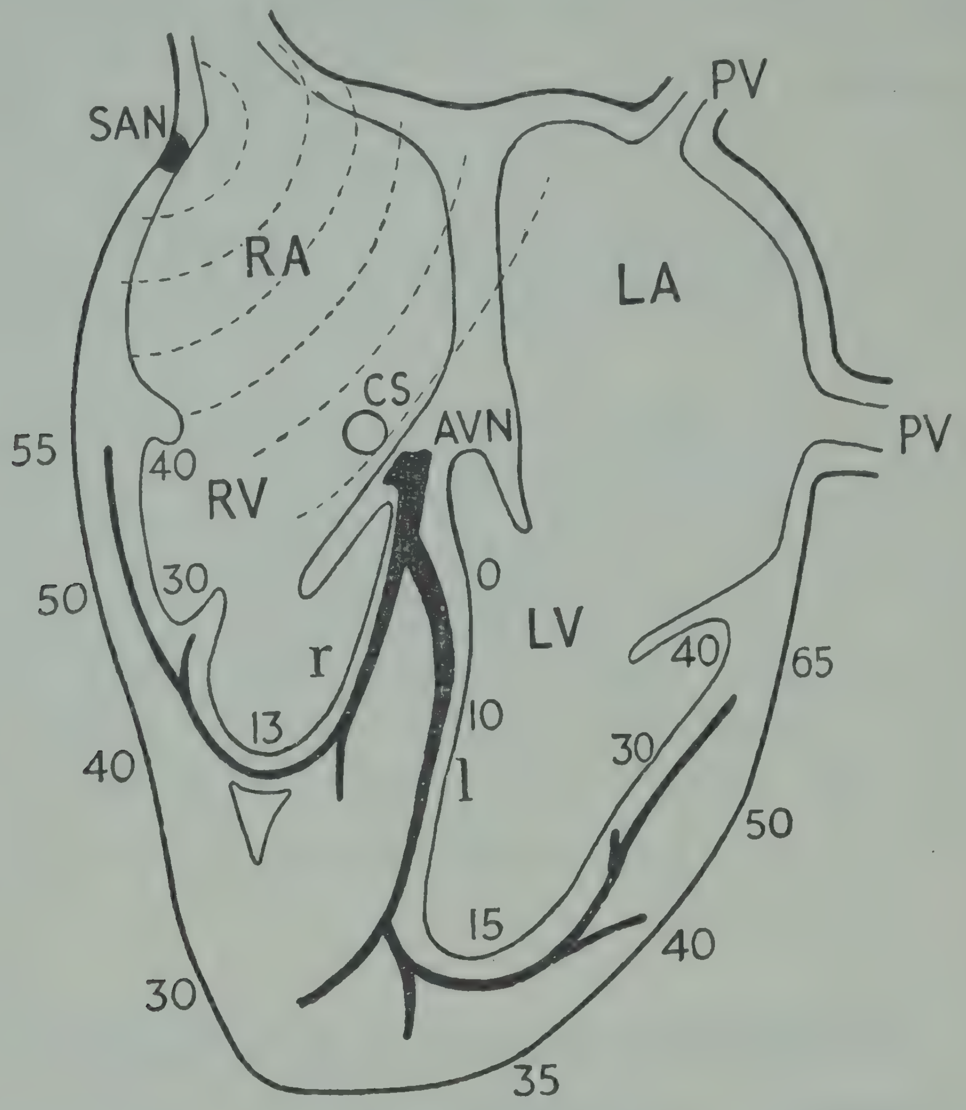


Fig. 28, 26

Diagrammatic section through the mammalian heart to illustrate the spread of the excitation process.

SAN = sinu-atrial node, AVN = atrio-ventricular node, RA = right atrium, LA = left atrium, RV = right ventricle, LV = left ventricle, CS = opening of coronary sinus, PV = pulmonary vein. Interrupted lines indicate the radiation of the excitatory process to all parts of the atria, the veins and, most important, to the AV node. Running down in the interventricular septum from the AVN is the main bundle of primitive conducting tissue (His), which divides into a main right branch (r) terminating in the right papillary muscle and base of the right ventricle and a main left branch (l) to the wall of the left ventricle. The figures 0 to 65 represent the time of arrival of the excitation process in msec reckoned from the time of arrival at the top of the interventricular septum. (*After* T. Lewis.)

Since the active portion of a muscle becomes electrically negative with respect to the inactive portion it is possible to explore the surface of the heart with electrodes connected to an oscillograph to find which part becomes active first. In this way it has been found that the excitatory process starts in the sinu-atrial node (S.A. node) lying on the right border of the opening of the superior vena cava and in the upper part of the sulcus terminalis which marks the junction between the embryonic sinus venosus and the atrium. From here the excitatory process and the wave of contraction spread in all directions (Fig. 28, 26), a little way along the superior and inferior venæ cavæ and over the atrial muscle fibres to the A.V. node in the atrial septum, whence the wave is carried along the A.V. bundle and through the Purkinje tissue to all parts of the ventricles.

The rate of conduction of the excitatory wave over the atria was estimated in the dog by Lewis as 0·8 and by Trautwein as 1·6 metres per second. Since the rate of spread was the same in all parts of the atrium it cannot possess any special conducting tissue. The excitatory wave passes more slowly through the A.V. node at a rate of not more than 0·2 metre per second. The delay between the times of atrial contraction and ventricular contraction is to a large extent due to the slow speed of conduction in the A.V. node since the A.V. bundle conveys the excitatory process at a speed of about 4 metres per second. The ordinary fibres of the ventricle, which are not specially adapted for conducting, but for contracting, are capable of carrying the excitatory wave at only 0·4 metre per second, that is, about a tenth of the rate in the conducting tissue. This accounts for the later arrival of the excitatory process on the pericardial surface of the heart as compared with the endocardial. Numerical values for the delay are given in Fig. 28, 26. The result of the greater speed of conduction in the specialized tissue is that the maximum difference in time between the beginning of the contraction of the first ventricular fibres and that of the last fibres to contract is only 0·07 seconds. The importance of this can be realized by considering that if there were no special conducting tissue this time would be increased to 0·7 seconds and the heart would be a very inefficient pump.

CARDIAC METABOLISM

The muscular tissue of the heart resembles skeletal muscle cells in containing glycogen, adenosinetriphosphate and phosphocreatine, but its metabolic behaviour is somewhat different from that of skeletal muscle. The Starling heart-lung preparation cannot be used to study cardiac metabolism since lactic acid is produced in large amounts in the lungs as the result of the breakdown of blood glucose. If the lungs are replaced by an artificial oxygenator this difficulty is overcome. Since the glycogen content of the heart in such a preparation remains fairly constant, it would appear either that glycogen is not utilized or that it is replaced as fast as it is used. The heart can take up glucose from the blood although not enough to supply its requirements, but it consumes lactic acid and pyruvic acid readily. Thus in the heart-oxygenator preparation lactic acid is used up at a rate of 200 mg./100 g./hour and glucose at a rate of only 70 mg./100 g./hour. The utilization of these two substances would involve an oxygen uptake of 200 ml. O_2/100 g./hour, whereas the actual amount used is 350 ml./100 g./hour. The difference appears to be due to the fact that cardiac muscle can utilize fat directly.

When skeletal muscle is poisoned with iodoacetic acid it is unable to form lactic acid but it can still contract until all the phosphocreatine it contains has been broken down (p. 775). On the other hand, cardiac muscle poisoned with iodoacetic acid continues to contract for an indefinite period provided that it is supplied with oxygen and lactate. The energy for contraction is obtained by the breakdown of phosphocreatine which in turn is resynthesized; the energy for this reaction is provided by the oxidation of lactic acid. If the oxygen supply is cut off lactic acid can no longer be oxidized and, since glycogen cannot be broken down to provide energy in the presence of iodoacetic acid, the phosphocreatine stores are soon exhausted and the heart very quickly fails.

From these and other observations it is clear that lactic acid is the main source of energy for cardiac contraction.

The rhythmic activity of the heart and the enzymic synthesis of acetylcholine are in some way linked together. A freshly excised heart beats vigorously (p. 431) and its power of synthesizing acetylcholine is high but when this falls the heart stops (see also Chapter 51).

Coronary Circulation

Blood is carried to the heart by the right and left coronary arteries (Fig. 28, 27), the first branches of the aorta (Fig. 28, 3), but there is considerable variation in the way that they are distributed to the various parts of the heart. The anastomoses between the terminal branches of the coronary arteries are confined to channels less than 40 μ wide, but narrowing of these arteries by disease leads to the development of anastomoses some 40 to 200 μ in size which may maintain a fully adequate blood flow. There are about 5000 capillaries per sq. mm of cardiac muscle, which is about twice that found by Krogh in skeletal muscle. Probably about three-fifths of the blood returning from the heart muscle flows into the right atrium by way of the coronary sinus, the remainder returning, as already described, by small channels opening directly into the various chambers.

The blood flow through the coronary circulation can now be measured in man and is found to be about 65 ml. per 100 g. of heart tissue per minute. This gives an approximate value for the outflow from the coronary sinus of 200 ml. per min, equivalent to about 4 per cent. of the cardiac output. On this basis, at a cardiac output of 30 litres per minute during severe exertion the volume of blood passing through the coronary circulation would be 1200 ml. per minute. The actual flow is less than this because the proportion of the total cardiac output that traverses the coronary circuit progressively diminishes as the output rises. The pressure within the coronary sinus can also be measured after the introduction into it of a catheter passed into the right atrium; it has been found that the pressure during systole is about 16 mm Hg and during diastole about 8 mm Hg.

The coronary blood flow depends very much on the arterial pressure, a fact which is to be expected on purely mechanical grounds. Thus when the arterial pressure rises, as it does in muscular exercise, more blood is forced through the coronary vessels and so the oxygen supply to the cardiac muscle is increased. Disproportion between the heart's demand for oxygen and the supply of oxygen through the coronary vasculature, that is to say relative ischæmia, leads to dilatation of the coronary vessels and an increase of blood flow of as much as 500 per cent. Since neither carbon dioxide nor lactic acid increase the flow by more than 50 per cent. the vasodilatation is attributed to anoxia which is probably the most important factor in regulating the blood flow to the needs of the myocardium. The coronary circulation is unusual in that the blood in the coronary sinus is almost

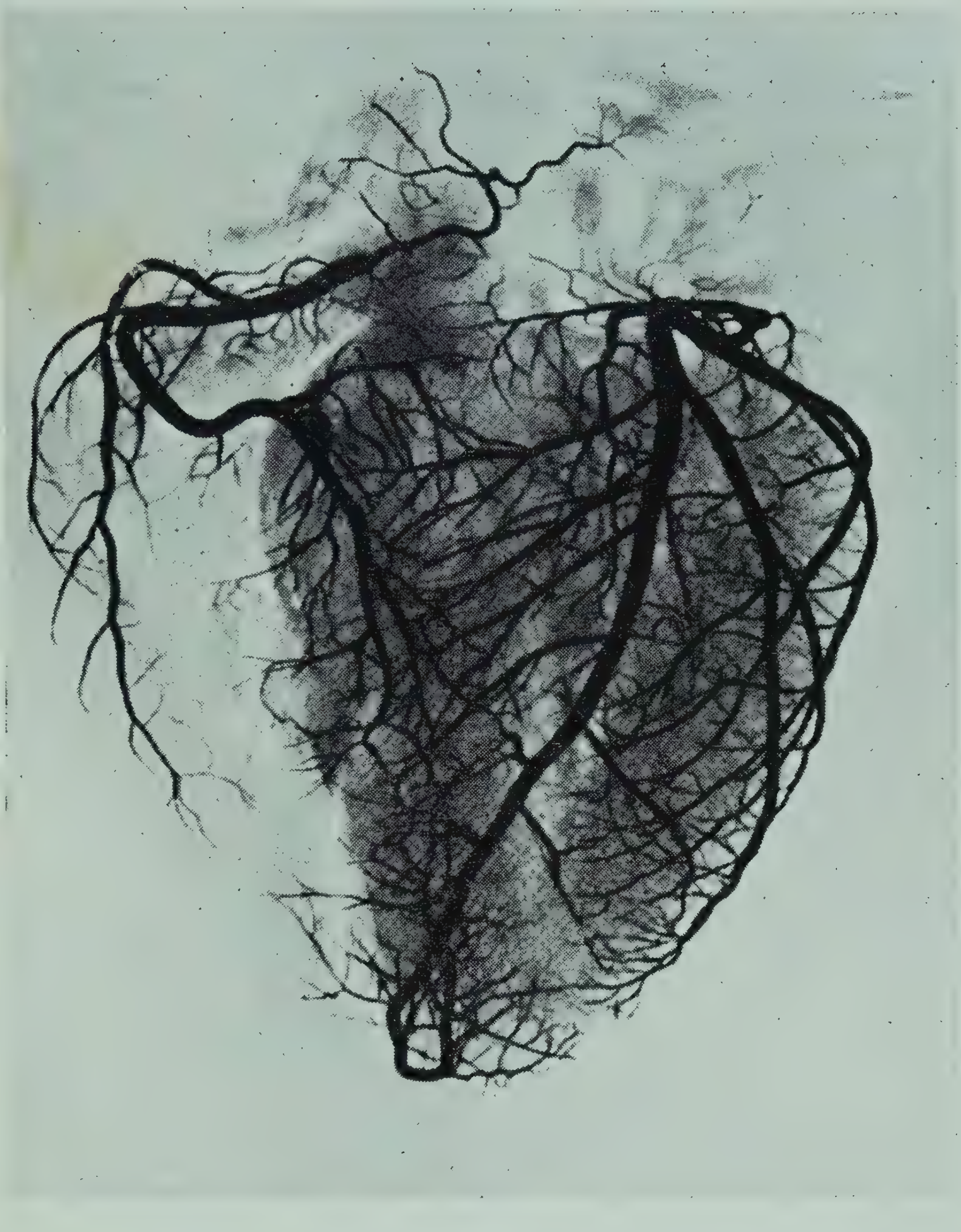


Fig. 28, 27

Coronary arterial system of a healthy heart from a man aged 18 years. X-ray taken after injection of the arteries with barium sulphate. The left coronary artery is considerably larger than the right; the large branch going to the apex is the interventricular (anterior descending) branch of the left coronary artery (J. Cowan & W. Ritchie (1935), *Diseases of the Heart*, 3rd ed. p. 9. London: Arnold).

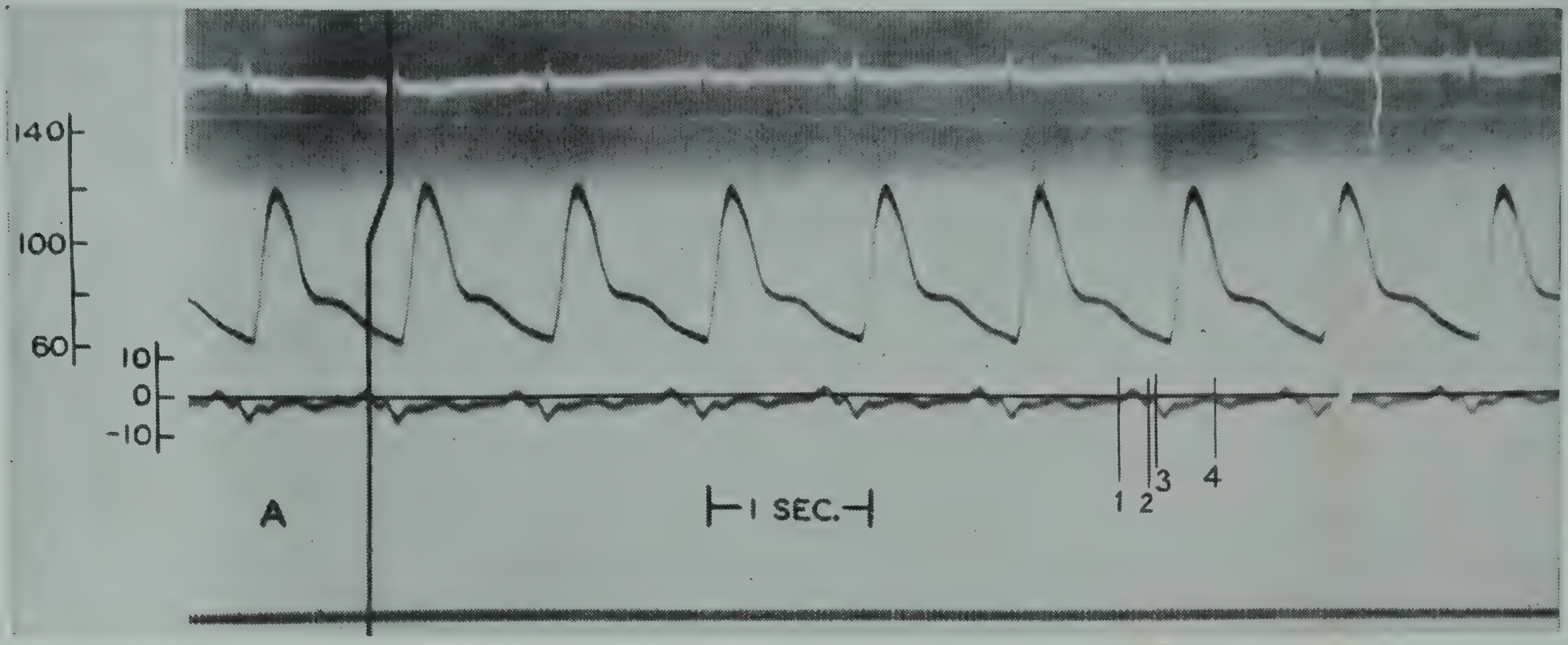


FIG. 28, 28

Upper trace, electrocardiogram. Middle trace, femoral arterial pressure (scale 60 to 140 mm Hg). Lower trace, right atrial pressure (scale −10 to +10 mm Hg). In lower trace: 1, beginning of atrial systole; 2, closure of AV valves; 3, descent of base of heart during ventricular ejection; 4, AV valves open. The vertical black line on the left is drawn through points in the three traces which correspond in time. Time is registered on the base line in divisions of 0·2 and 0·04 sec. (R. A. Bloomfield, H. D. Lanson, A. Cournand, E. S. Breed & D. W. Richards (1946), *J. clin. Invest.* **25**, 643).

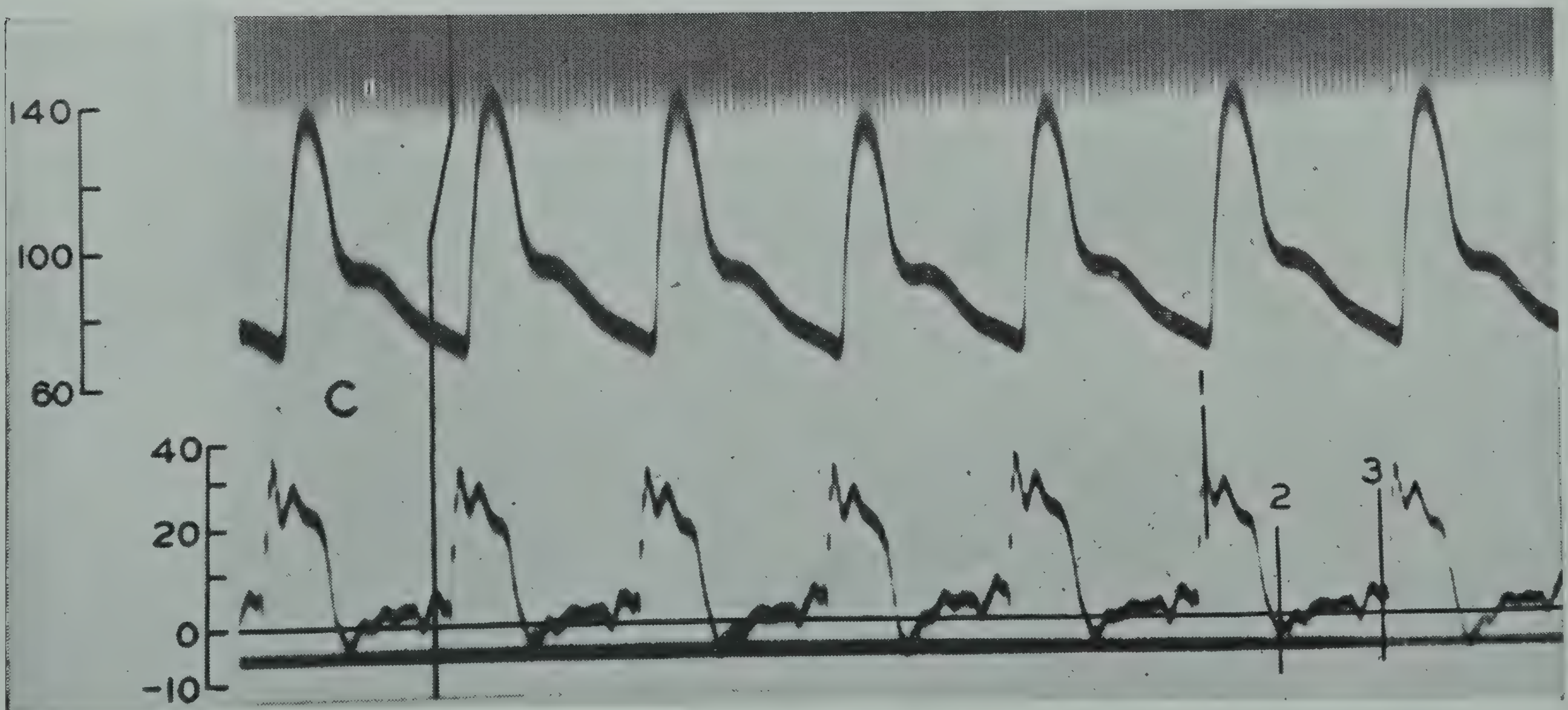


FIG. 28, 29

Upper trace, femoral arterial pressure (scale 60 to 140 mm Hg).
Lower trace, pressure in right ventricle (scale −10 to +40 mm Hg).
In lower trace, 1 = maximum systolic pressure, 2 = diastolic pressure early in diastole, 3 = diastolic pressure late in diastole. The small positive wave immediately preceding 3 represents atrial systole. Time is marked on the upper edge of the figure in 0·2 and 0·04 sec. The vertical black line is drawn through points in the two traces which correspond in time (R. A. Bloomfield, H. D. Lanson, A. Cournand, E. S. Breed & D. W. Richards (1946), *J. clin. Invest.* **25**, 644).

completely desaturated containing only 5 to 9 ml. of oxygen per 100 ml. of blood. The supply of oxygen to the heart can therefore be augmented only by an increase in coronary flow.

The coronary blood vessels are richly supplied with sympathetic and parasympathetic nerves which, however, are probably of little importance in regulating coronary blood flow. Section of the vagus nerves leads to an increased heart rate which is accompanied by dilatation of the coronary arteries. Stimulation of the sympathetic nerves or the addition of adrenaline to the circulating blood similarly produces cardiac acceleration and coronary dilatation. It is difficult to separate experimentally the effects on coronary flow of changes in aortic pressure from those produced directly through the action of the coronary vasomotor nerves, but it seems quite clear that, in the dog, alterations in the tonic action of the vagus nerves produce variations in coronary blood flow, inhibition of vagal tone leading to vasodilatation. Finally it is not always possible to exclude the influence of metabolites produced in the heart muscle itself. It is likely that hydrostatic and chemical factors are pre-eminent in governing the blood supply to cardiac muscle.

It will be obvious that since the coronary circulation must provide blood to a mass of muscle contracting, say, 70 times a minute there must be peculiarities of inflow and outflow not seen in other vascular beds. During ventricular systole blood is squeezed out of the capillaries so that the curve of outflow follows closely the curve of intraventricular pressure. Although the coronary arteries are compressed during a contraction so that flow is impeded or even arrested, as much as 30 to 40 per cent. of the total flow may occur during systole.

Since part of the output of the left ventricle flows into the coronary circulation during diastole the flow of blood along the aorta does not represent the total cardiac output. The total output is given by the flow along the pulmonary artery.

NERVOUS REGULATION OF THE HEART

The heart-lung preparation shows that even the isolated heart has considerable powers of regulating its activity according to need. This is also true for the denervated heart *in situ*. A dog whose cardiac nerves have been cut can survive in fairly good health for several months, although incapable of severe exercise. In the intact animal, however, the activity of the heart is regulated to a large extent through the central nervous system.

The heart receives outgoing branches of the vagus nerves from the medulla, and sympathetic outgoing nerve fibres from the upper thoracic region of the spinal cord—Fig. 28, 30. The former come from the dorsal nucleus of the vagus, the autonomic vagal nucleus, and travel down the vagus trunk in the neck, passing by way of the superficial and deep cardiac plexuses to the heart. The fibres from the right vagus end mainly round the S.A. node and those from the left chiefly round the A.V. node. In these positions they arborize with ganglion cells and new postganglionic fibres pass in the coronary plexuses to the atria, to the A.V. bundle and to the base of the ventricles. The sympathetic fibres arise chiefly from the second and third thoracic segments of the spinal cord and pass in the anterior roots as myelinated nerve fibres to the cervical sympathetic ganglia, where they form synapses. From here postganglionic nonmyelinated nerve fibres pass to the superficial and deep cardiac plexuses, where they intermingle with the fibres of the vagus. The postganglionic sympathetic fibres pass without interruption to the muscle fibres of the heart.

Although there is no doubt much individual variation in the way the autonomic fibres are distributed to the heart, it is probable that this description applies in general to man.

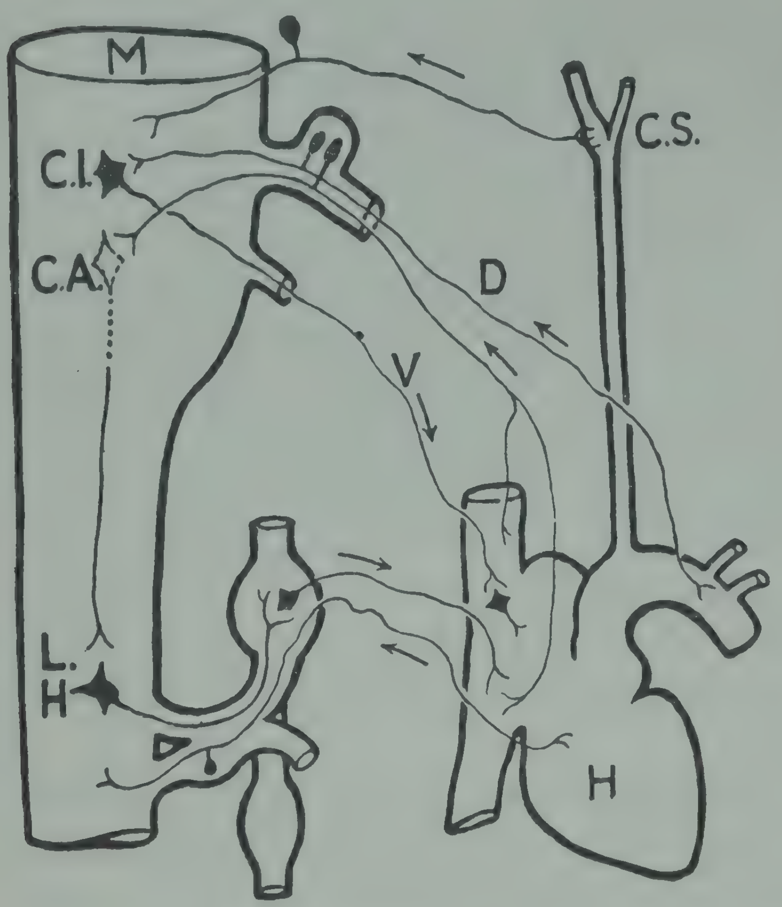


Fig. 28, 30

Simplified diagram of the nerve supply of the heart

C.I. = cardio-inhibitory centre.
C.A. = cardio-accelerator centre.
M = medulla oblongata.
L.H. = lateral horn of spinal cord.
D = depressor fibre (afferent vagal).
V = efferent vagal fibre.
C.S. = carotid sinus.
H = heart.

The Vagus Nerve.—If the vagus nerve is divided in an animal, such as a dog or a cat, and the peripheral end stimulated it will be found that after a latent period of less than a second the heart beats more slowly and the blood pressure falls considerably. The diastolic interval is prolonged, ventricular filling is greater and the stroke volume is increased (Fig. 28, 33). When the stimulation is stronger the heart may be brought to a stop in diastole but it will beat again after the stimulation ceases. Even if stimulation is continued the ventricles (but not the atria) will, after an interval, resume contractions. This is called *vagus escape*, and indicates that the ventricles, when deprived of impulses from the atria, can beat independently at their own rhythm. This escape also suggests that the vagus has little direct effect on ventricular activity, although there are other factors operating when the heart is slowed. For example blood accumulates on the right side, calling forth the right atrial reflex (p. 466). The fall of aortic pressure may also bring the aortic and carotid sinus reflexes (p. 462) into action. The main effect of stimulation of the right vagus, distributed chiefly to the S.A. node (the pacemaker), is a reduction of the atrial and therefore the ventricular rate, and stimulation of the left vagus depresses mainly the A.V. conducting tissue. Most parts of the heart, however, receive fibres from both nerves and stimulation of either the right or the left vagus produces similar effects, namely slowing, diminution of strength and duration of systole and an increase in the time taken by the excitatory process to pass from the S.A. to the A.V. node and along the A.V. bundle.

Impulses are continually passing down the vagus nerves to exert a retarding action in the heart and it is chiefly by variations in this *vagal tone*, as it is called, that alterations in heart rate are produced. If the vagus nerves to the heart of an animal are cut (Fig. 28, 35), or if in man an injection of atropine is given to prevent the effect of the vagal impulses on the heart, the heart rate rises (Fig. 28, 31). The effect of vagal section is most marked in animals capable of periods of sustained exertion (such as the horse and the hare) which normally have a slow resting pulse (Table 28, 32). In athletes the heart is often slow (35 to 40 beats per min.) and it may be that the tonic effect of the vagus is more marked in them than in non-athletic persons.

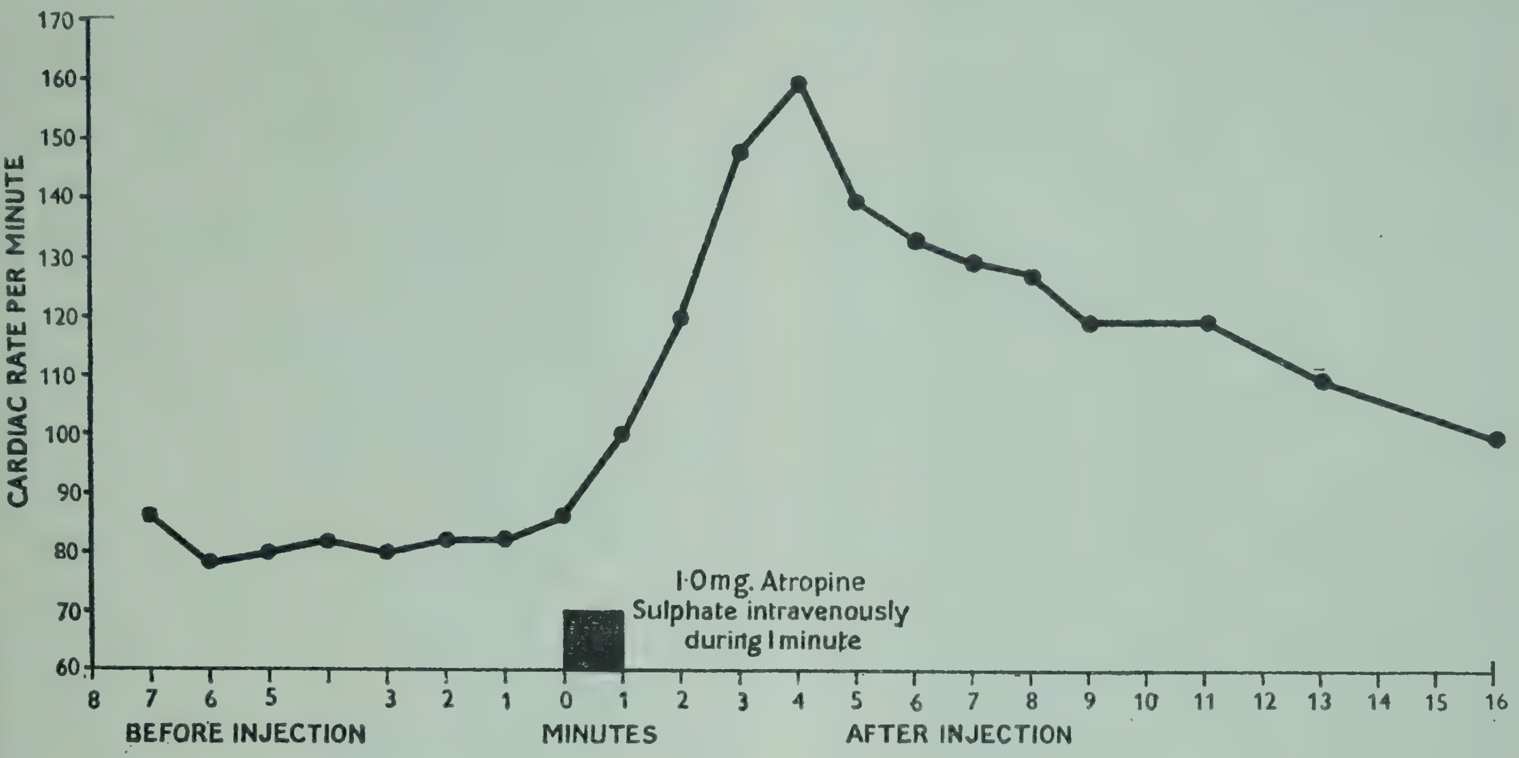


Fig. 28, 31

Healthy male aged 32 years. Increase in cardiac rate after atropine. The response to a given dose of atropine is very variable. Large doses are necessary to obtain complete vagus release.

Table 28, 32

	Rabbit	Hare
Pulse rate at rest .	205	60 to 70
Pulse rate after vagal section . . .	321	264

(From A. J. Clark (1927), *Comparative Physiology of the Heart*, p. 58. Cambridge University Press.)

It was the study of the action of the vagus nerve on the heart that led Loewi to suppose that when this nerve is stimulated small amounts of a highly active material are liberated at its endings. This is now known to be acetylcholine. Perfusion of a frog's heart with acetylcholine slows the heart just as vagal stimulation does. If the perfusate from one frog's heart is used to perfuse a second then it can be shown that the acetylcholine produced by stimulation of the vagus nerve of the first heart passes over in the perfusing fluid to the second heart, which is also slowed. Further details of the production of acetylcholine by nerves are given on p. 788.

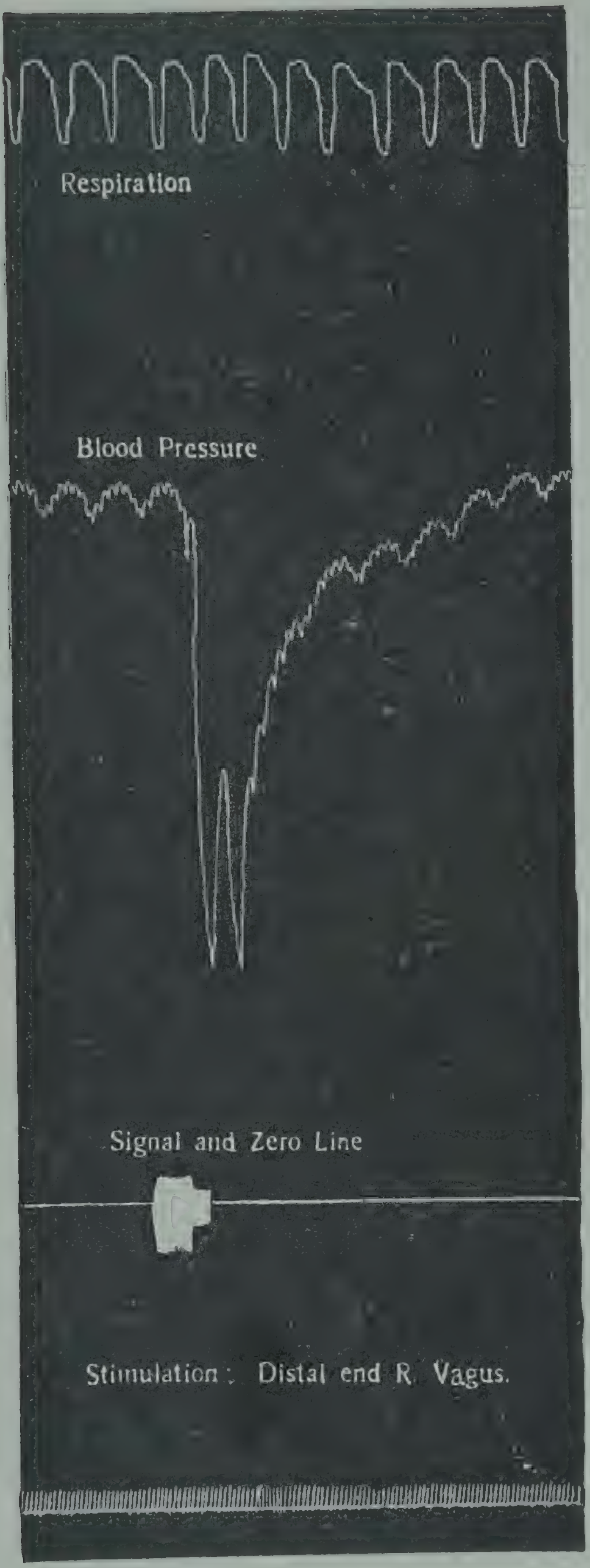


FIG. 28, 33

The effect of stimulation of the distal end of the right vagus of a dog on blood pressure and respiration. Time in seconds.

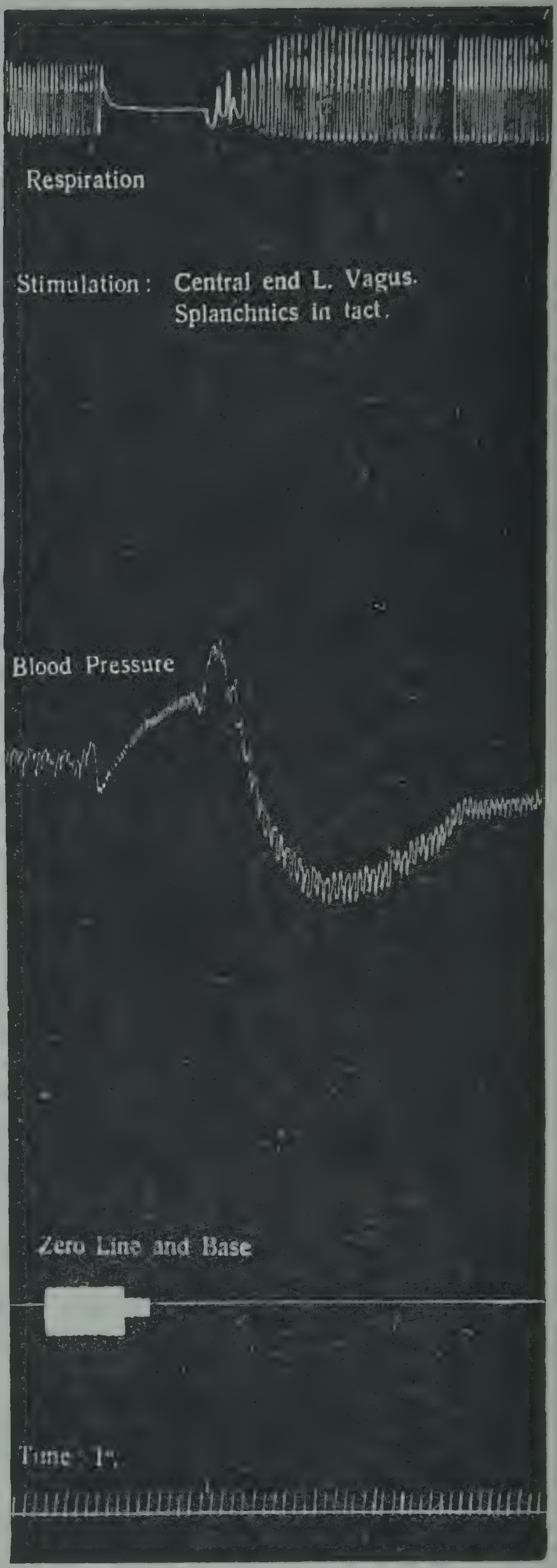


FIG. 28, 34

The effect of stimulation of the central end of the left vagus nerve of a dog. The right vagus and the splanchnics were intact. An upward movement of the respiratory trace indicates inspiration. Time in seconds. Note the effects on the blood pressure and respiration.

(*By courtesy of E. W. H. Cruickshank.*)

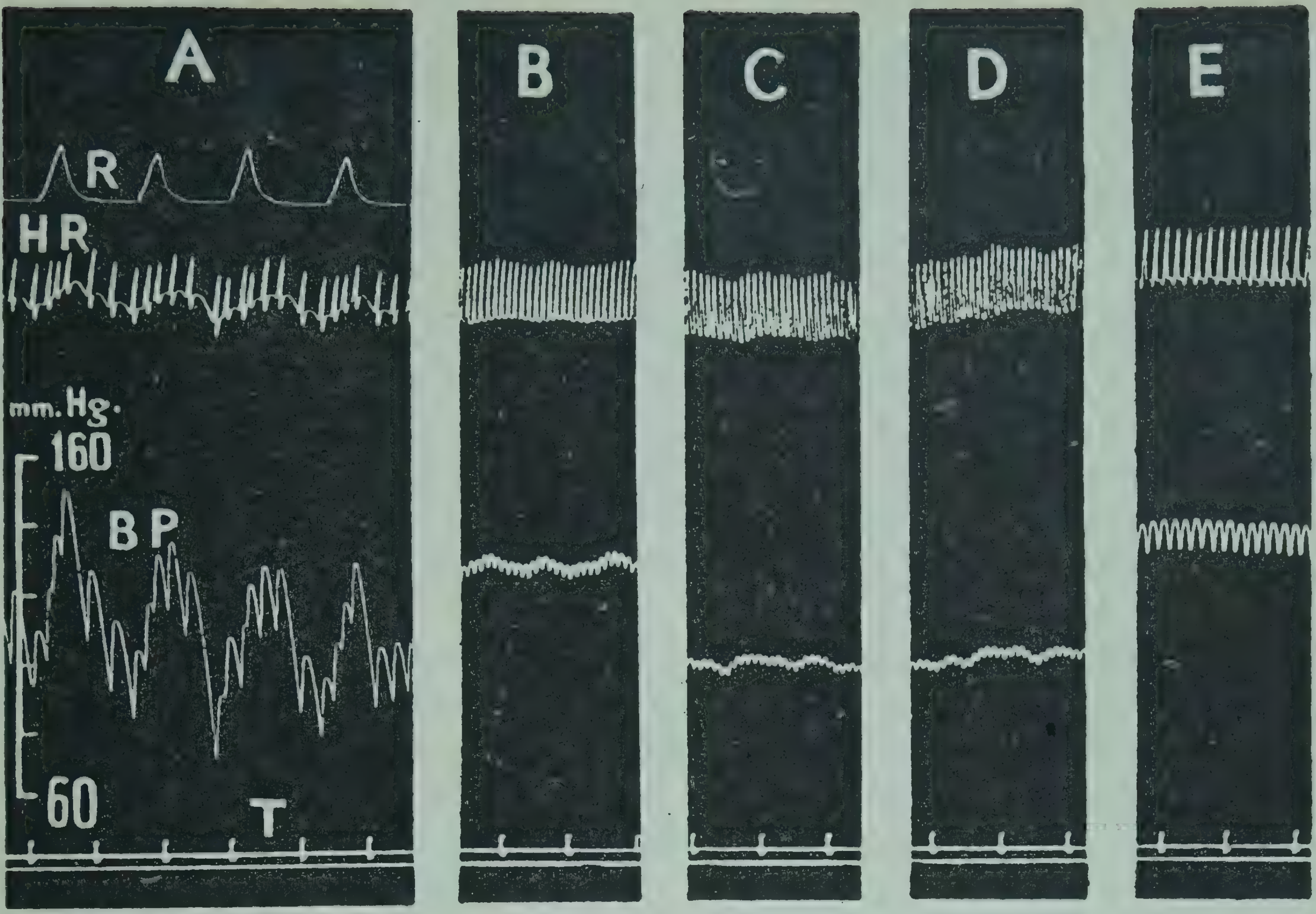


FIG. 28, 35

The heart rate during the process of cardiac denervation. Dog: 14 kg.; chloralosane. A, normal. B, after cutting the cervical vagi, artificial respiration on. C, after removal of the left thoracic sympathetic chain. D, after excision of the right thoracic chain excepting the stellate ganglion. E, after extirpating the right stellate ganglion. R denotes respiratory movements recorded with stethograph; HR is the heart rate taken with Hürthle's manometer; BP is blood-pressure record with mercury manometer; T is time record in 3 sec. (Adli Samaan (1935), The Antagonistic Cardiac Nerves and Heart Rate, *J. Physiol.* **83**, 333).

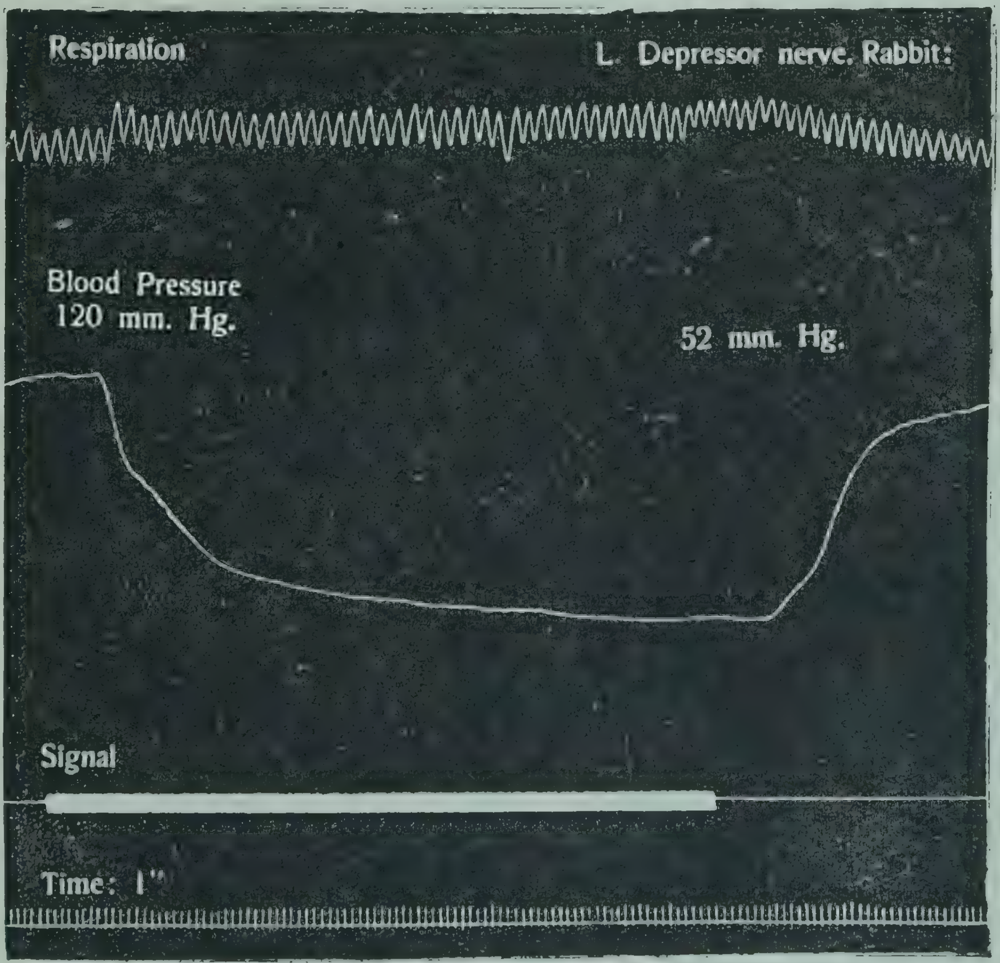


FIG. 28, 36

The effect of stimulation of the central end of a depressor nerve in a rabbit (p. 463). There is a marked fall of blood pressure with very little change in respiration.

(*By courtesy of E. W. H. Cruickshank.*)

Atropine prevents both these effects, not by stopping the production of acetylcholine but by preventing in some way its action on the cells. Normally acetylcholine is broken down by the enzyme cholinesterase to acetic acid and choline both of which are inactive. This cholinesterase action can be prevented by eserine and in consequence eserine itself slows the heart.

The Sympathetic Nerves.—Stimulation of the sympathetic nerves affects mainly the S.A. node. After a latent period of six to ten seconds there is an increase in the rate of the heart at the expense of a shortening of the diastolic interval. There is also an increase in the force of the contraction. The sympathetic nerves appear (like the vagus nerves) to be constantly transmitting impulses to the heart because their section is followed by a slight slowing (Fig. 28, 35).

It seems likely that the material formed at the nerve endings when the sympathetic nerves are stimulated contains adrenaline and noradrenaline. The injection of adrenaline into the intact animal produces an acceleration and augmentation of the heart beat just as sympathetic stimulation does. The injection of noradrenaline into the intact animal however, while it causes an augmentation of the heart beat, leads to a reduction of rate, possibly reflexly through the vagus. When adrenaline or noradrenaline is added to the blood of a heart-lung preparation the ventricle beats very forcibly, the durations of both systole and diastole are reduced, the heart volume diminishes, the venous pressure falls, the arterial pressure rises, the oxygen usage increases and the coronary vessels dilate.

It has been shown by cardiac catheterization in the human subject that slow intravenous infusion of small doses of adrenaline (for example 3 μg. per min), insufficient to cause a rise of heart rate or to raise the blood pressure, causes a rise of cardiac output (Fig. 28, 37).

It has recently been found that on stimulation of the sympathetic nerves to the heart of the dog noradrenaline, with little or no adrenaline, is liberated into the blood.

Afferent Nerves and Cardiac Reflexes.—Since the efferent vagal fibres arise in the dorsal nucleus of the vagus, and since afferent fibres of the vagus end in the same region, it is convenient to speak of this area of the medulla oblongata as the cardio-inhibitory centre (Fig. 28, 30). The vasomotor centre controlling the calibre of the blood vessels is also situated in this part of the brain and the two centres have numerous interconnecting fibres.

It has not been possible to define exactly the position of the cardio-accelerator centre which may lie near the cardio-inhibitory centre in the medulla or perhaps in the hypothalamus. Nerve fibres pass down from the centre to the cells in the intermediolateral tract of the grey matter of the spinal cord from which arise the sympathetic cardiac nerves.

The two chief afferent nerves concerned in the regulation of cardiac activity are (1) those which arise in the heart and aorta and pass up the vagus, and (2) those which arise in the carotid sinus and pass by the glossopharyngeal nerve to the medulla oblongata (Fig. 28, 30). In addition, myelinated nerve fibres pass from the heart by various routes through the inferior cervical and the upper four thoracic ganglia and by the upper four thoracic posterior nerve roots into the spinal cord. These afferents probably subserve pain sensation.

Cardio-aortic Nerve.—The main ingoing or afferent nerve, the cardio-aortic nerve, arises from endings in the left ventricle and aortic arch. In man and in most animals the fibres ascend in the trunk of the vagus, but in the rabbit they form a separate nerve bundle called the depressor nerve. If the depressor nerve of a rabbit is cut and the peripheral end stimulated there is no effect, but stimulation of the central end causes a reflex slowing of the heart and a fall of blood pressure (Fig. 28, 36). In animals in which the cardio-aortic fibres are carried in the vagus trunk, stimulation of the central end of the cut vagus similarly produces cardiac slowing and a fall of blood pressure (Fig. 28, 34). The afferent impulses travel in the vagus to the cardio-inhibitory centre and efferent impulses pass down the other intact vagus nerve and produce cardiac slowing. If both vagus nerves are cut the heart is not slowed when the central end of one vagus is stimulated but there is usually a fall of blood pressure due to a dilatation of the abdominal vessels supplied by the splanchnic nerves.

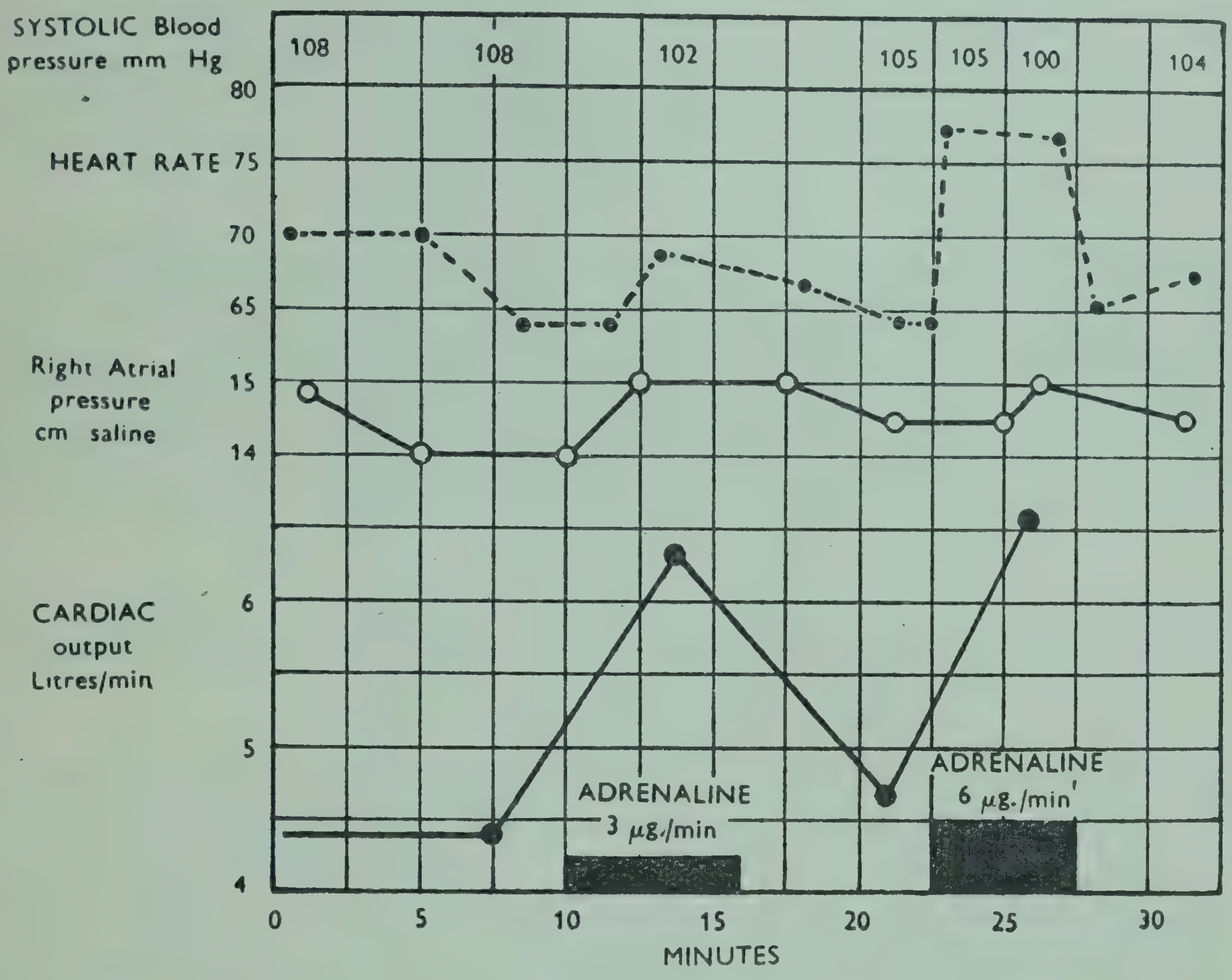


FIG. 28, 37

Effect of adrenaline infusion at 3 μg. a minute. Cardiac output increases without change in heart rate or right atrial pressure or rise in blood pressure. When the dose of adrenaline was doubled, the heart rate increased (J. McMichael & E. P. Sharpey-Schafer (1944), *Brit. Heart J.* **6**, 38).

The cardio-aortic nerve endings are stimulated when the pressure in the left ventricle and aorta is high, that is, they are *baroceptors.* In the dog these endings are stimulated so long as the pressure is above 60 mm Hg. If a record of the nerve impulses is made at the same time as a record of the blood pressure it is found that a burst of impulses passes up the cardio-aortic nerve at each systole and a similar burst passes down the vagus nerve to the heart. The higher the aortic pressure the greater is the frequency of these impulses. In cats, and probably in other animals, there are baroceptors in the myocardium, stimulation of which results in a fall of arterial pressure and heart rate. The afferent fibres run in the vagus nerves.

Carotid Sinus Nerve.—The dilatation at the commencement of the internal carotid artery called the carotid sinus contains nerve endings which are sensitive to stretching like those of the cardio-aortic nerve. The fibres pass to the medulla in a branch of the glossopharyngeal nerve. Electrical stimulation of the carotid sinus or its nerve, or distension of the sinus by increase of the pressure within it, produce a reflex slowing of the heart and a fall of blood pressure exactly as described in the case of the cardio-aortic nerve. In fact, the cardio-aortic and the carotid sinus nerves should be regarded as two afferents of one regulating mechanism for preventing any undue rise of blood pressure and for avoiding an excessive load being taken by the heart. For this reason these are sometimes called 'buffer' nerves. The effect of stimulation of the endings in man can be shown by digital pressure over the bifurcation of the carotid artery lying behind the medial edge of the sternomastoid muscle at the level of the hyoid bone (Fig. 28, 39). Such pressure will often produce some slowing of the heart beat even in normal individuals but the effect is more striking if the beat is rapid. Carotid sinus pressure is of particular value in averting certain kinds of paroxysmal tachycardia which occur in disease (Fig. 28, 38). In 1863, long before these reflexes were understood, Marey enunciated his 'law' which states that *pulse rate varies inversely as arterial pressure.* Although this generalization is not always true of the intact animal the reflex basis for it is now clear.

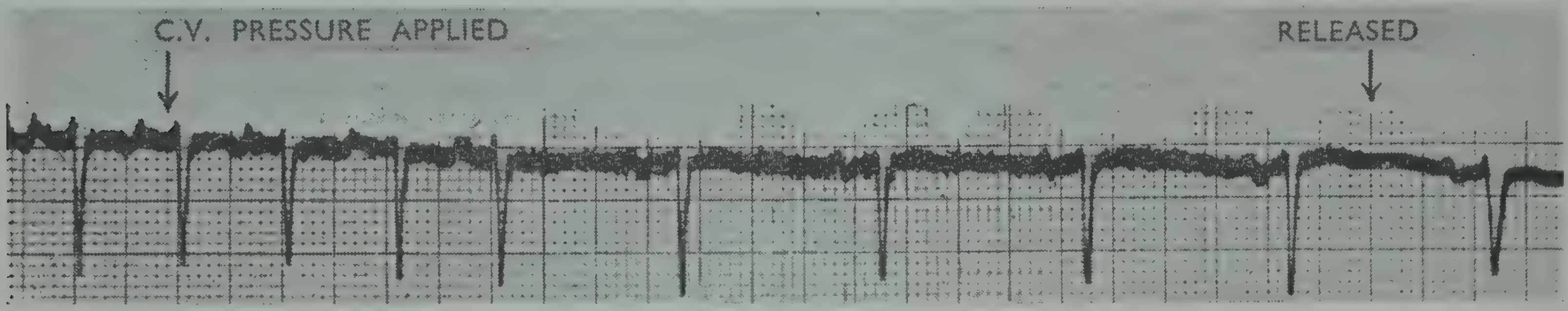


FIG. 28, 38

Effect of left carotid pressure on heart rate in a patient suffering from atrial flutter (p. 477). The rate at the beginning of the trace is about 150/min and at the end 75/min. Time in $\frac{1}{5}$ and $\frac{1}{25}$ sec.

The reflex effects upon the heart of alterations of pressure within the carotid sinus are very readily demonstrated. While a blood-pressure record is being made from a dog or cat by the method to be described later, the carotid sinus pressure can be reduced temporarily to nearly zero by holding the intact common carotid artery tightly between the fingers. There follows an immediate reflex increase in the rate of the heart with a rise of blood pressure. These purely nervous phenomena may be reinforced by a reflex secretion of adrenaline from the suprarenal medulla. When the fingers are released, allowing the pressure in the carotid sinus to increase again, there is an immediate fall in the heart rate and in the systemic pressure. Heymans has demonstrated these effects more elegantly by isolating the carotid sinus of a dog from the rest of the circulation, leaving the carotid sinus nerve intact and perfusing the sinus by a pump, or by connecting it to the arterial system of another dog. When the pressure in the

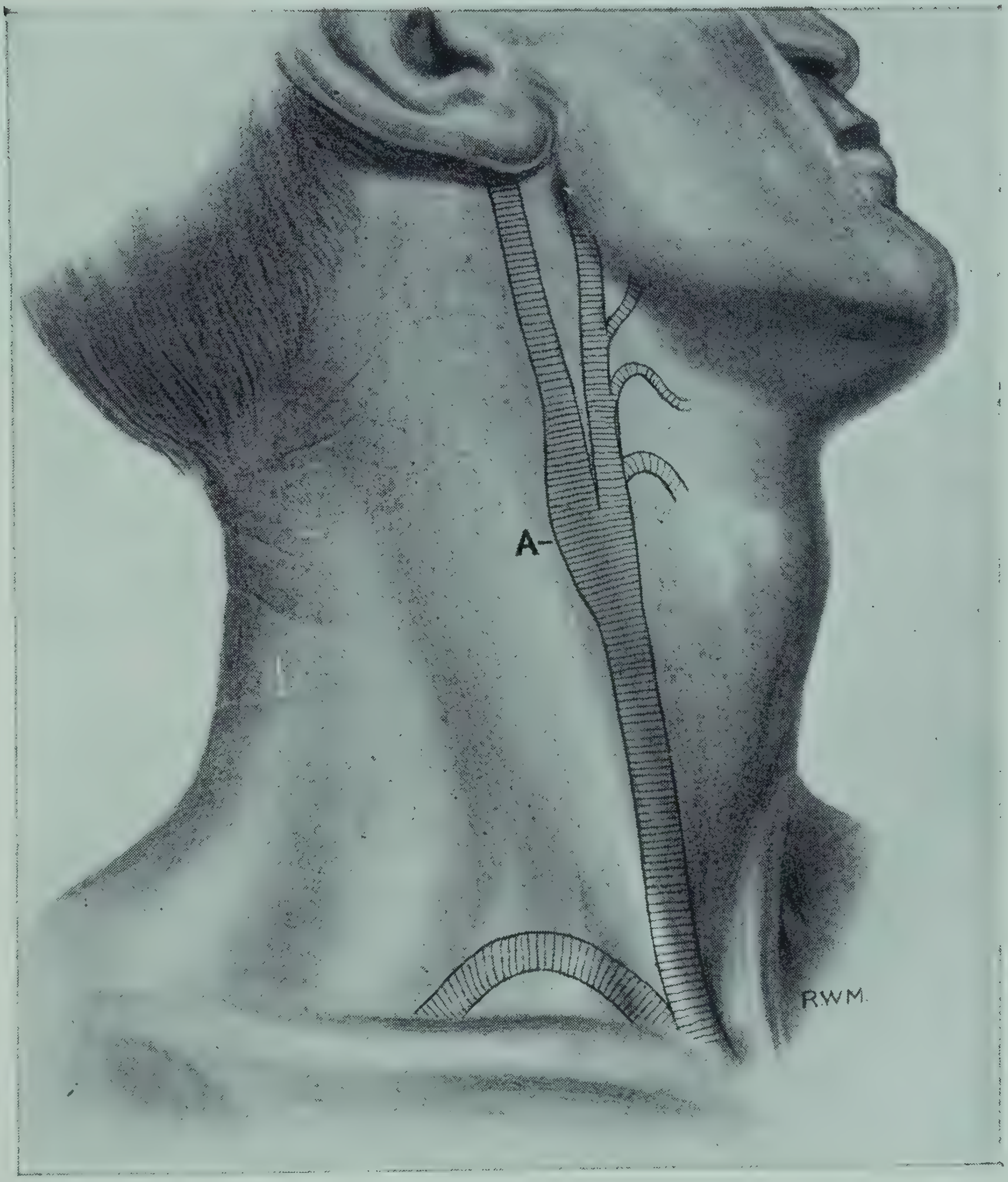


FIG. 28, 39

Diagram showing the position of the carotid sinus (A) in the neck.

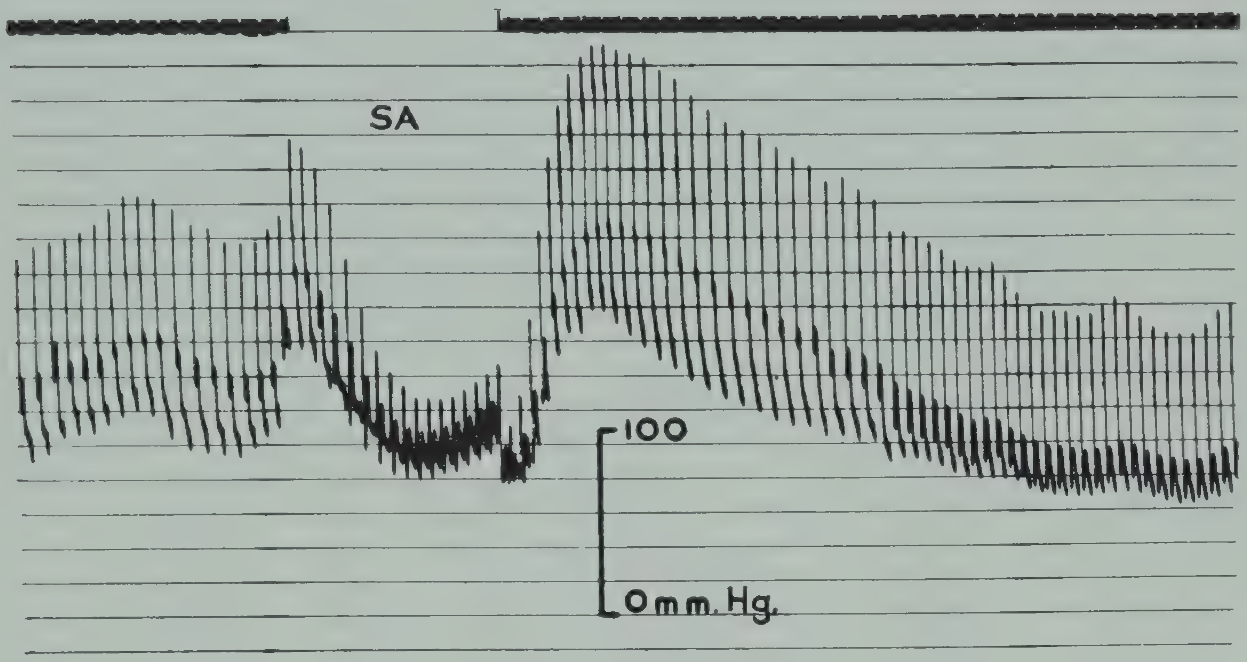


FIG. 28, 40

Continuous record of blood pressure in the brachial artery, made by a capacitance manometer, during the performance of the Valsalva manœuvre (p. 436) at the break in the time signal (upper black line). Time in seconds. (P. Howard, G. L. Leathart, A. C. Dornhorst & E. P. Sharpey-Schafer (1951), *Brit. med. J.* ii, 382.)

isolated carotid sinus is raised a little then the systemic pressure falls a little and the heart is slowed; a further rise of carotid sinus pressure is followed by a further fall of systemic blood pressure and so on. When the carotid sinus pressure is reduced to its original value the heart accelerates and the systemic blood pressure rises.

Before the carotid sinus mechanism was recognized it was known that a rise of arterial pressure in the head resulted in a fall of heart rate and blood pressure in the remainder of the body. This was assumed to be a direct effect of the increased pressure on the medullary centres. By experiments very similar to those detailed above it has been shown that a rise of arterial pressure in a head isolated from the body, but connected to it by the vagus nerves, results in a fall of blood pressure in the body when the carotid sinus is intact but not after its nerve is cut. This means that the effect of increase of pressure in the head is a reflex originating in the carotid sinus and not the result of direct stimulation of medullary centres.

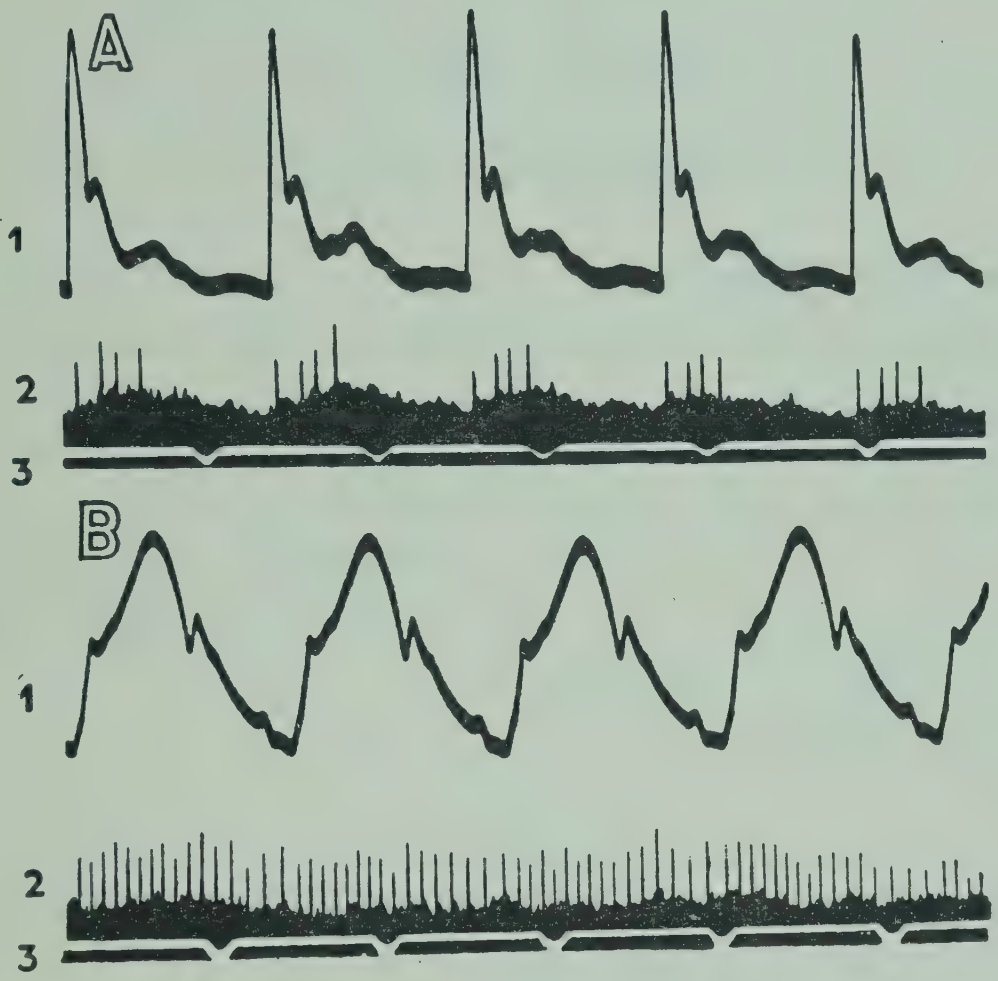


FIG. 28, 41

In each tracing curve 1 shows the arterial pressure registered by a membrane manometer; curve 2 shows the discharge from a single end-organ in the carotid sinus of a rabbit; curve 3 gives the time in fifths of a second. In A, B.P. = 55 mm Hg. In B, B.P. = 135 mm Hg (D. W. Bronk & G. J. Stella (1932), *J. Cell. comp. Physiol.* **1**, 113).

The impulses in the carotid sinus nerve are shown beautifully in the tracing of Bronk and Stella from a single nerve ending (Fig. 28, 41). It can be seen that the number of afferent impulses passing along the nerve is proportional to the carotid sinus pressure and that during diastole, when the pressure is falling, the frequency of the impulses is less.

The Carotid Body.—The carotid body situated behind the bifurcation of the common carotid artery contains blood spaces (sinusoids) and cells ('glomus cells') richly supplied with afferent nerve fibres which join the carotid sinus nerve. The nerve endings are sensitive to alterations in the composition of the blood in the sinusoids and are, therefore, called *chemoreceptors.* The carotid body is chiefly concerned with regulation of respiratory movements, but an increase of carbon dioxide, or a decrease of oxygen tension, causes an acceleration of the heart as well as alterations of respiration (p. 563) and vascular tone. The effects on the heart are of little physiological importance except perhaps when the blood pressure is low (p. 508 and p. 513).

TABLE 28, 42

FACTORS WHICH CAUSE

ACCELERATION OF HEART	SLOWING OF HEART
Rise of right atrial and central venous pressure: right atrial reflex	Fall of right atrial and central venous pressure
Fall of B.P. in aorta and carotid sinuses: (*a*) cardio-aortic reflexes (*b*) reflex secretion of adrenaline and noradrenaline	Rise of B.P. in aorta and carotid sinuses: (*a*) cardio-aortic reflexes
Painful stimuli	
Rise in body temperature	Fall in body temperature
Decrease in O_2 tension in carotid body	
Increase in CO_2 tension in carotid body	
Emotion or excitement	Emotion
Exercise	Sleep
Increased metabolism	Decreased metabolism

Reflexes arising from the Great Veins and Right Atrium.—Bainbridge in 1915 produced in animals acceleration of the heart by increasing the pressure in the right atrium and great veins by large infusions of saline or blood. He concluded that the increase in cardiac rate was the result of a reflex initiated by the increase in pressure. In cats there are afferent fibres in the vagus nerves whose frequency of discharge is proportional to the pressure in the great veins but it is doubtful if the Bainbridge reflex is mediated through such fibres. The part played by reflexes arising from the venous side of the heart in the regulation of the cardiac rate is not at present known.

During exercise the heart rate and blood pressure remain elevated above the resting level in spite of the carotid sinus and aortic depressor reflexes. It must, therefore, be assumed that the pressor mechanisms take precedence over the depressor reflexes. It will be recalled that cardiac acceleration can also occur as a result of a fall of arterial pressure which reduces the number of impulses passing up to the cardio-inhibitory centre.

THE RATE OF THE HEART

The heart rate counted under basal conditions in young healthy adult males lies between 60 and 65 beats per minute, the figures for women being somewhat higher. However, under ordinary conditions the heart rates of men average about 78, and of women 84 beats per min and the American Heart Association accept between 50 and 100 beats per minute as the normal range. There is a tendency for the heart rate to fall as age advances (Fig. 28, 43). The cardiac rate increases progressively with increasing exercise up to some 200 beats per min but the rate of increase is less in trained subjects. The rate at which the cardiac frequency returns after exercise to the pre-exercise level is used as an index of physical fitness, the initial rate ideally being restored within 2 minutes of the end of exercise.

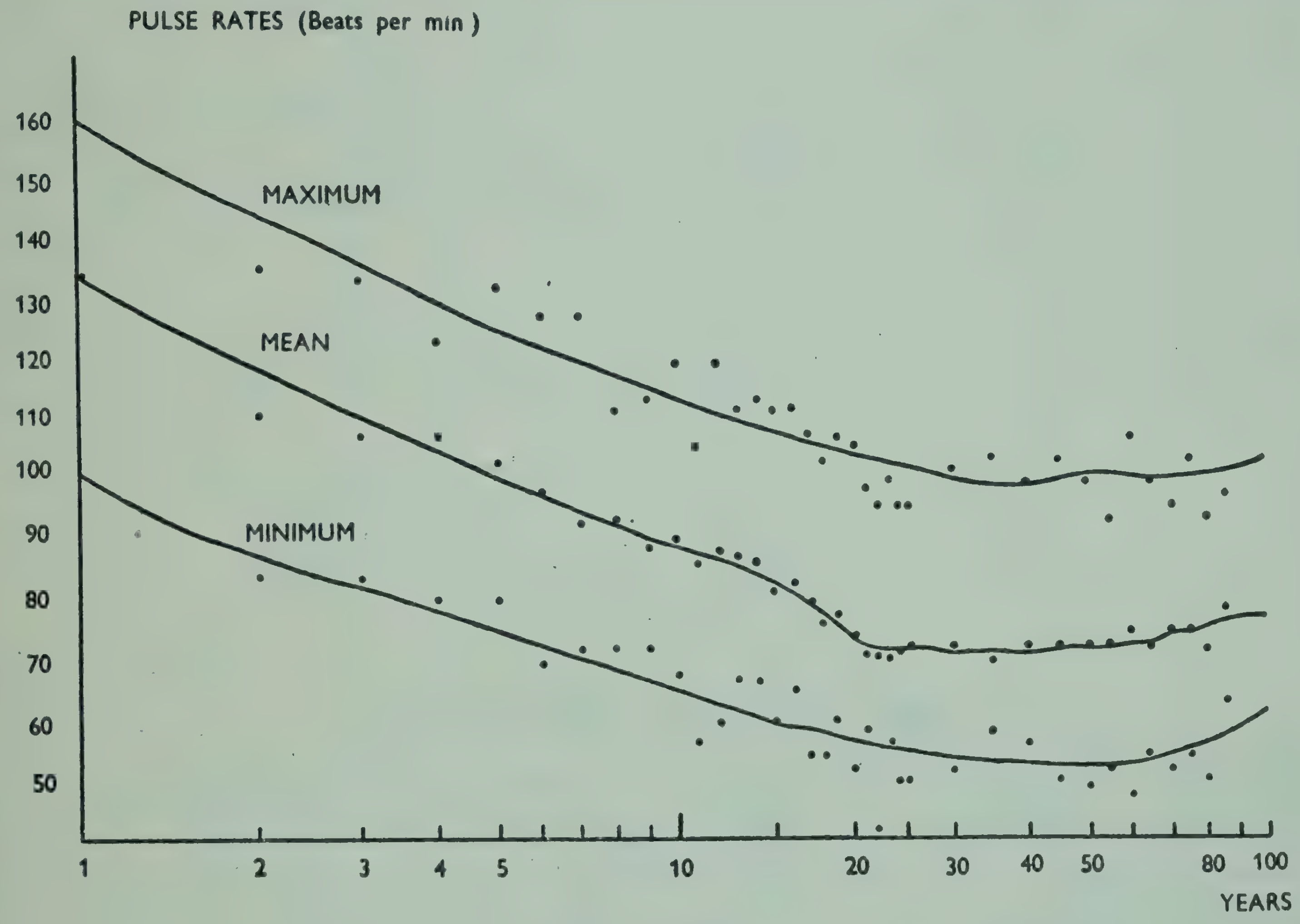


Fig. 28, 43

Diagram to show variation of resting pulse rate with age constructed from data of Tigerstedt, *Lehrbuch der Physiologie des Kreislaufes.*

The cardiac rate is elevated by all kinds of emotional disturbances (Fig. 28, 44) and, conversely, it falls progressively during sleep. The difference between the waking and sleeping pulse rates may therefore be of some significance as an index of emotional tension. The heart rate is affected by body temperature and observations on a large number of individuals has shown that the heart rate can be predicted with reasonable accuracy by the formulæ

$$11 \times {}^{\circ}\mathrm{F} - 1018 = \text{heart rate or } 20 \times {}^{\circ}\mathrm{C} - 674 = \text{heart rate.}$$

In other words, the heart rate increases on the average by 11 beats per minute for each °F rise in temperature (20 beats per min per °C). Some of the factors determining frequency of the heart beat are summarized in Table 28, 42.

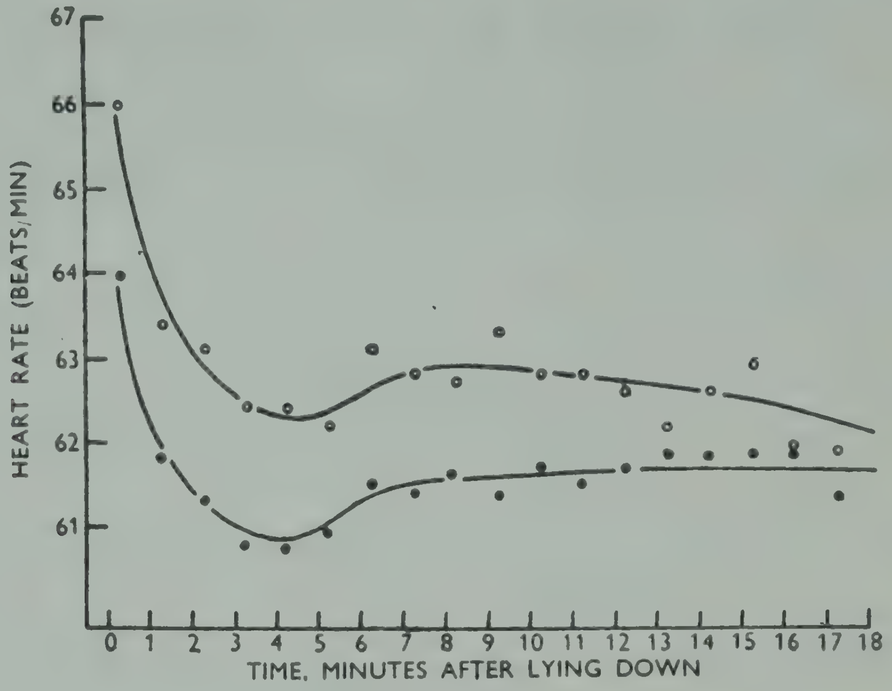


Fig. 28, 44

Mean heart rate of forty-six healthy young men for 18 min after lying down. ○——○, on first visit to laboratory; ●——●, on second visit, approximately 8 days later. (J. M. Tanner (1951), *J. Physiol.* **115,** 391–409.

THE ELECTROCARDIOGRAM

The history of electrocardiography clearly shows how the development of physiology depends on technical advances. Köllicker and Müller in 1856 laid the nerve of a sciatic-gastrocnemius preparation on a frog's heart and noted that the muscle twitched with each beat of the heart. The capillary electrometer was first used in 1876 to detect cardiac action potentials and with this instrument the first human electrocardiogram was obtained by Waller in 1889. The first accurate records were obtained by the string galvanometer invented by Einthoven in 1901 (Fig. 28, 45) which is still in regular use to-day. Since the advent of the thermionic valve amplifier much less sensitive but more robust galvanometers can be used, such as the moving-iron oscillograph devised by Matthews in 1928 and the cathode-ray oscillograph.

Each cycle of cardiac activity is accompanied by electrical potential changes which can be detected at the surface of the body and a record of such alterations in electrical potential is called an electrocardiogram. It is not possible, however, to determine from the record the actual potentials in the cardiac muscle because the heart lies in an irregularly shaped body containing organs which differ in their electrical properties; for example the conductivity of liver and chest wall is relatively greater than that of the lungs. The interpretation of the electrocardiogram is, therefore, to some extent empirical, that is, it must be based on experience and experiment.

Theoretical Derivation.—The general form of the electrocardiogram is derived by considering the series of events described in Fig. 28, 46. At rest each heart cell is bounded by a polarized membrane having negative ions on the inside and an equal number of positive ions on the outside. The heart, because of the intimate connexions between its cells, can be regarded as a large syncytium in the middle of a mass of tissue. In the resting state, that is during diastole, as at Fig. 28, 46, *a*, no current flows and the galvanometer is not deflected, so that its record on a moving surface is a horizontal line. When the cell is activated the polarized membrane is broken down as in Fig. 28, 46, *b*. The electrode to the right is still connected to the layer of positive ions on the surface, while the electrode on the left is now connected to the negatively charged interior of the cell. The galvanometer is therefore deflected and an upright wave is recorded. When all parts of the cell are depolarized (Fig. 28, 46, *c*) the galvanometer returns to zero. As recovery occurs the membrane becomes repolarized (Fig. 28, 46, *d*) and current flows in the opposite direction; the record falls below the isoelectric line, returning to it once more when repolarization is complete (Fig. 28, 46, *e*). The two parts of this biphasic wave, though opposite in sign, are not, as might be expected, equal in size and similar in shape. In the first place the process of repolarization is a slower one than the initial depolarization, and secondly the order in which the cells of the heart are depolarized depends on the spread of the excitatory process from cell to cell, whereas recovery depends on the activity of the individual cells. Thus some parts of the heart are more quickly restored to the resting state than others. For these two reasons the second part of the diphasic wave is smaller in amplitude and different in form from the first part. This theoretical derivation should be regarded as a much simplified account of the actual events. For example, it omits all reference to the reversal of membrane potential which occurs in active nerve and muscle cells (Chapter 41).

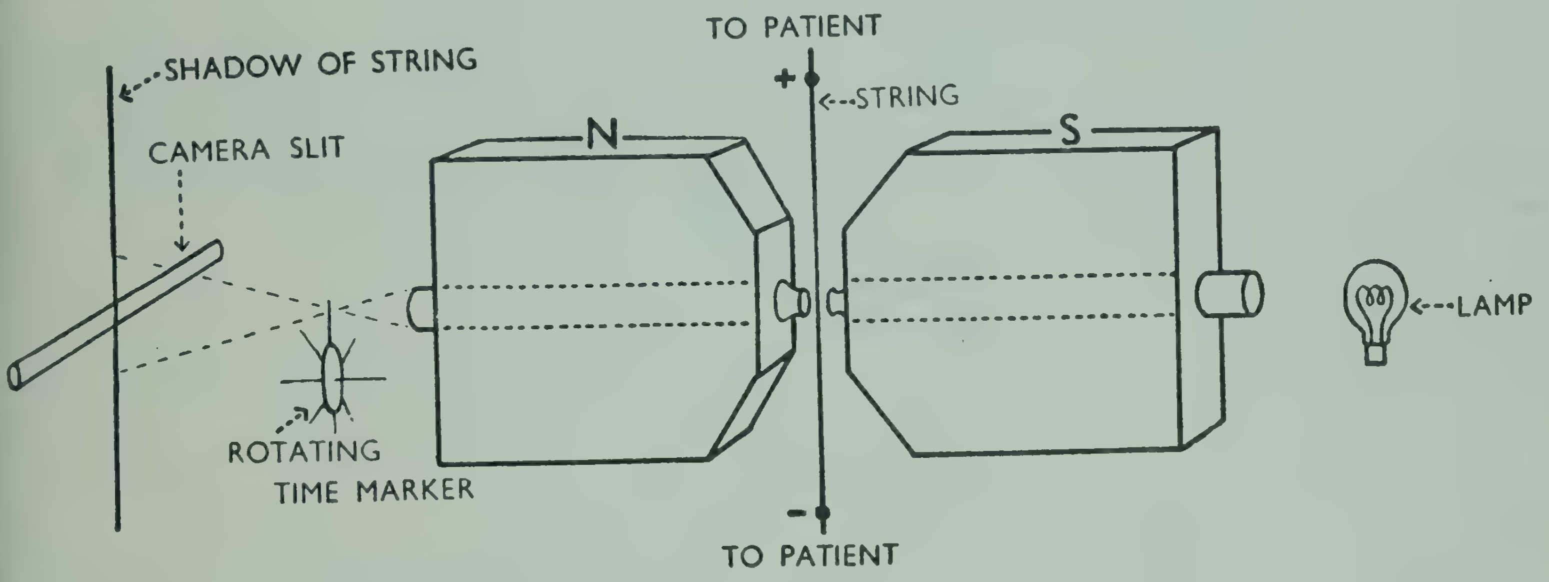


FIG. 28, 45

Einthoven's string galvanometer. The string is a gold-plated quartz fibre about 5 μ in diameter suspended in a powerful magnetic field. With the arrangement pictured in the diagram a positive potential at the upper end of the string will cause this to be deflected towards the reader but, because the direction of movement is inverted in the optical system, the image of the string moves from left to right on the camera slit. The rotating time-marker carries spokes which interrupt the light beam and cast regularly-spaced shadows on the slit.

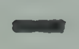
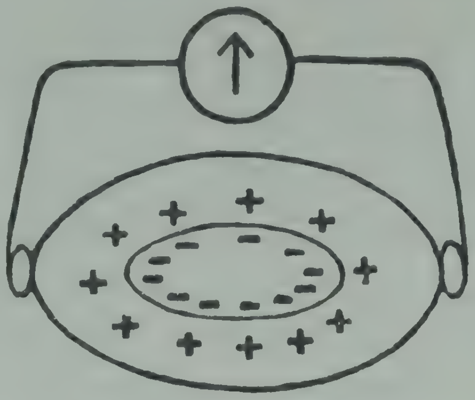
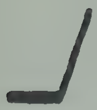
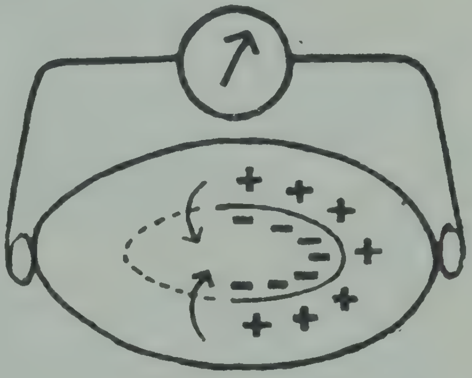
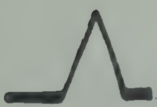
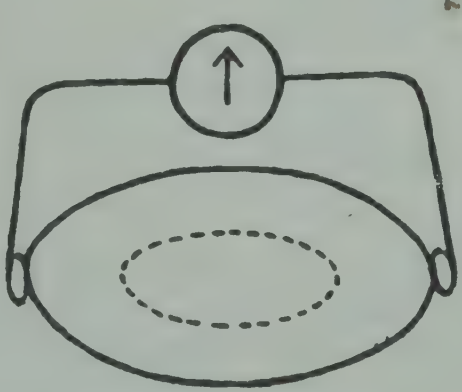
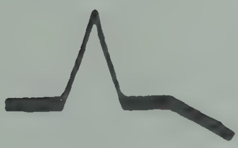
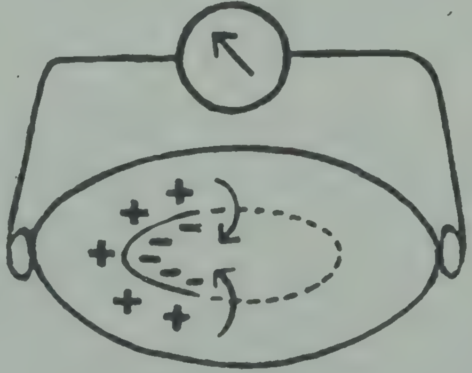
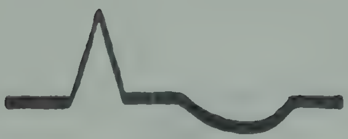
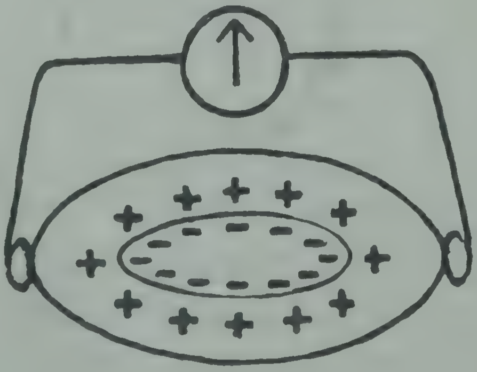
Oscillograph Record	Cardiac Events	Remarks
		*a.* At rest. Charged membrane, equal numbers of positive and negative ions. No difference of potential at surface.
		*b.* At the beginning of activation part of the cell is depolarized. There will be a flow of ions as indicated by the arrows and a potential is set up across the electrodes.
		*c.* When all parts are depolarized there is no difference of potential.
		*d.* Beginning of restitution of depolarized area.
		*e.* Restitution or repolarization complete. No flow of current.

Fig. 28, 46

Theoretical derivation of the electrocardiographic deflexions. (*After* Katz.)

The electrocardiogram is essentially the time record of electrical processes in the heart and the relation between these and other events in the cardiac cycle is given in Fig. 28, 9 which shows that electrical activity precedes mechanical activity by a short interval. The relation to the cardiac sounds is given in Fig. 28, 49. The locus of origin of the excitatory process and its speed of travel through the heart can be obtained from the electrocardiogram, which in special circumstances may also provide us with information about the state of the cardiac muscle. The electrical record does not, however, indicate the vigour of the heart, its efficiency, or its ability to perform the work demanded of it.

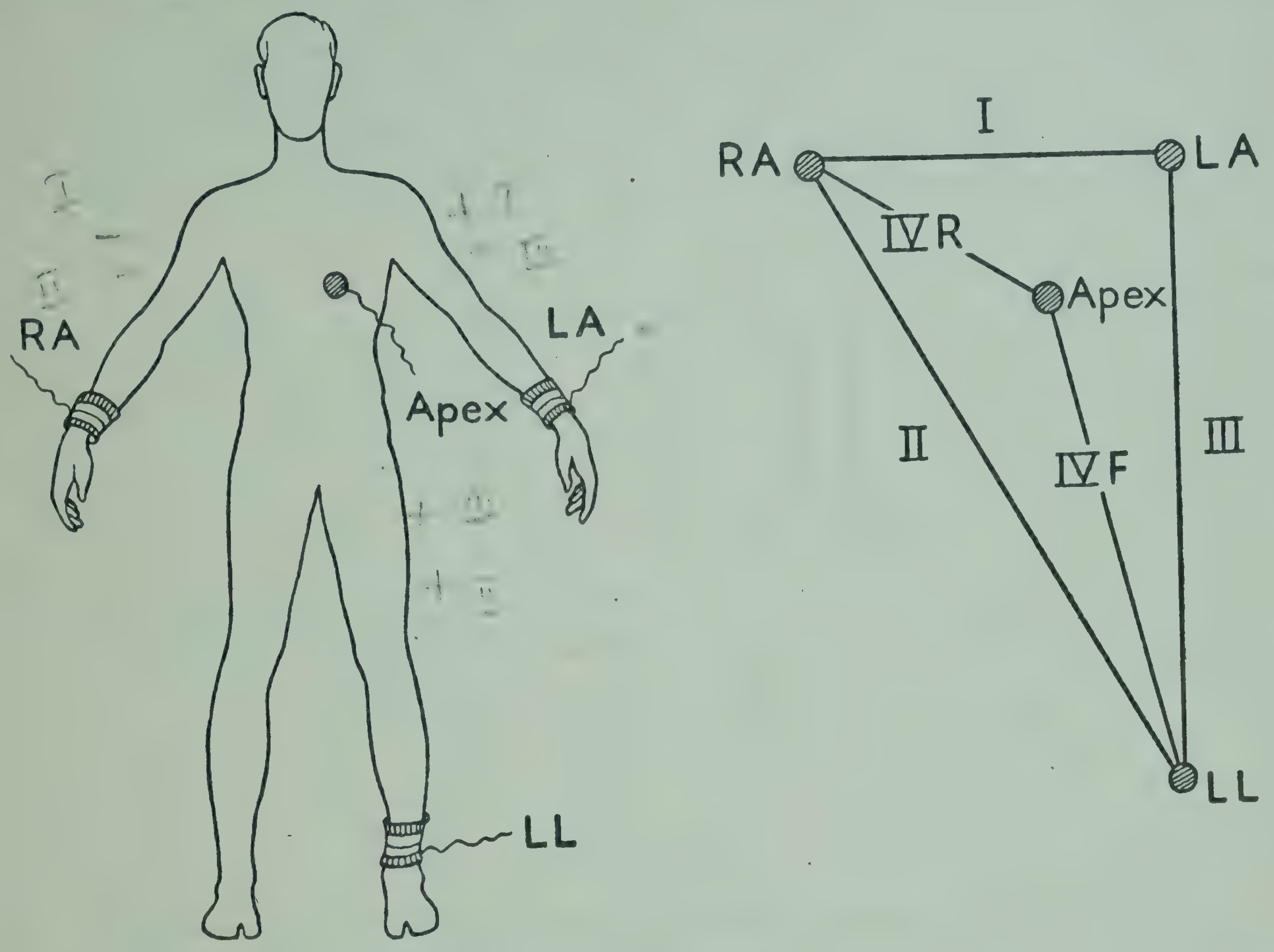


Fig. 28, 47

Leads used in clinical electrocardiography.

Lead I, right arm to left arm.
,, II, right arm to left leg.
,, III, left arm to left leg.
,, IV R, apex to right arm.
,, IV F, apex to left leg.
(A = arm; L = leg; R = right; L = left.)

A small area of skin is first rubbed with an abrasive jelly and then covered with a piece of thin lead sheet which is bandaged in position: a wire connected to the sheet goes to the galvanometer. The apical electrode, about 1 inch in diameter, is a flat cylinder of metal. The diagram to the right gives a scheme of the connexions for the standard leads.

The Human Electrocardiogram.—For clinical purposes the recording galvanometer is connected to the right and left arms and left leg of the subject, or to the chest over the apex beat as shown in Fig. 28, 47. Leads from other præcordial areas are now found to be of great value in the diagnosis and localization of coronary disease. Because the electrical resistance of the dry skin surface is considerable it is desirable to rub the skin gently with glass-paper, or with abrasive jelly containing powdered silica, to reduce the resistance between the body and the metal electrode. In the case of the string galvanometer, which has a relatively low resistance (several thousand ohms), polarization in the skin can distort the record seriously. It is advisable, therefore, to connect the patient to such an instrument by non-polarizable electrodes. The effect of polarization is negligible in the case of valve amplifiers in which the input resistance is usually of the order of half a million ohms.

A normal electrocardiogram is represented diagrammatically in Fig. 28, 48, which is enlarged for ease in labelling. The standard photographic record is taken with the apparatus adjusted so that a vertical movement of 1 cm indicates a potential change of one millivolt (mV).

The P wave (Fig. 28, 48) begins as the excitatory process passes from the S.A. node to the atria, its duration indicating the time taken for the process to spread throughout the atrial muscle. The QRS complex indicates the beginning of depolarization of the ventricles and it ends when all the ventricular muscle has been depolarized. The T wave signals the restitution (repolarization) process in the ventricles and the U wave, by no means always clearly distinguishable, is said to be due to alteration of the position of the heart in diastole. The PR interval, measured from the beginning of P to the beginning of Q, gives the time taken by the excitatory process to travel over the atria to the A.V. node and along the conducting tissue to the ventricular muscle. During the ST segment all parts of the ventricles have been depolarized; the record is, however, not quite isoelectric (horizontal) but inclined slightly in the same direction as the T wave, because restitution actually begins very slowly during this period. Restitution in the ventricles is completed during the T wave. The wave due to restitution in the atria is small and is obscured by the ventricular complex. In the TP interval, that is between two cycles of cardiac activity, the whole heart is in the resting polarized state and the record is

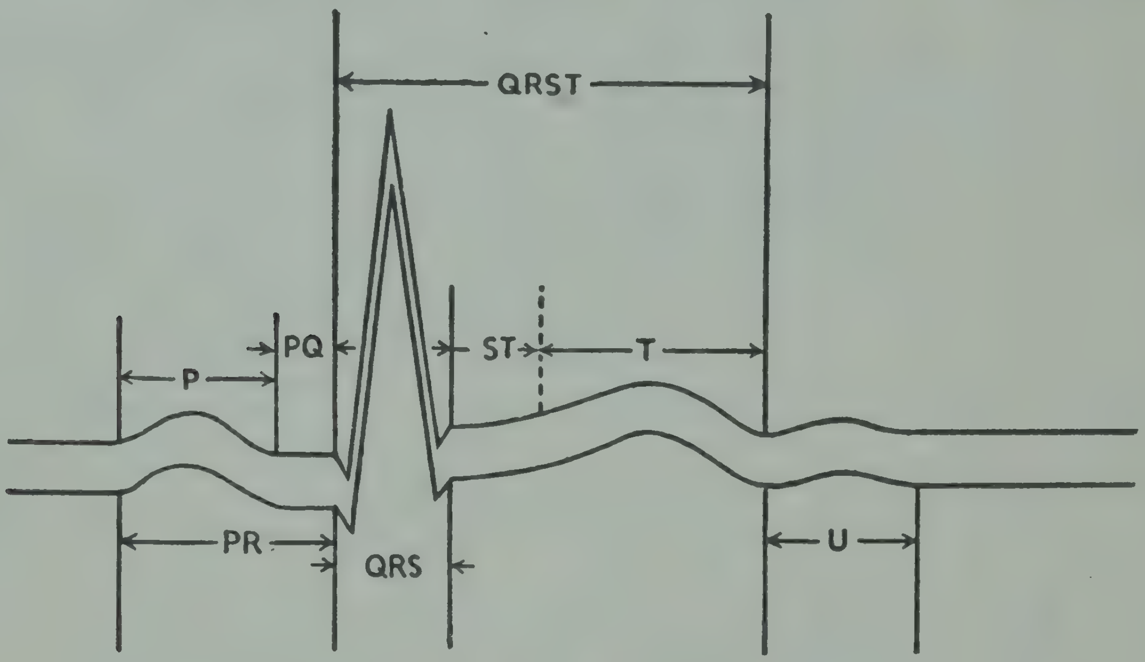


FIG. 28, 48

A diagram showing the beginning and end of the various parts of the electrocardiogram, namely the P wave, QRS complex, T and U waves, the P–R and QRST intervals, and the P–Q and S–T segments. The T–P segment from the end of T to the beginning of P is not labelled. By an agreed convention an upward deflexion in the finished record indicates electrical negativity of RA in leads I and II and of LA in lead III. In lead IV an upward deflexion indicates relative positivity of the apex (L. N. Katz (1946), *Electrocardiography*. 2nd ed. London: Kimpton).

isoelectric. In the diagram of the electrical changes in cardiac muscle (Fig. 28, 46) the T wave is opposite in sign to the QRS wave, whereas in the human electrocardiogram they are usually of the same sign. The explanation seems to be that the processes of the depolarization and repolarization follow different pathways and that repolarization is slower at the base than at the apex of the ventricles. These differences in path and time are sufficient to account for the upright T wave.

The P wave lasts from 0·06 to 0·11 sec. The PR interval is one of the most important measurements in the electrocardiogram; it usually lasts from 0·12 to 0·21 sec and varies inversely with the heart rate. An increase beyond 0·21 sec is almost always due to disturbance of conduction in the A.V. bundle. The QRS complex varies considerably in form, lasting from 0·06 to 0·33 sec. Values greater than this indicate either that the electrical events are prolonged in one ventricle or that the ventricles are not being excited simultaneously. The amplitude of the QRS complex is usually from ½ to 1½ millivolts in leads I, II and III, but may be up to 4 millivolts in lead IV. Although such data may be used to describe the normal electrocardiogram it must be emphasized that many types of so-called abnormality have been at one time or another described in records obtained from apparently normal hearts.

Clinical Applications.—The electrocardiograph was first successfully used in the clarification of disorders of rhythm. The following examples will suffice to show how it is used for this purpose. In sinus arrhythmia (p. 495 and Fig. 29, 22) the heart is activated from the pacemaker. The excitatory process travels over the muscle and conducting tissue normally so that the electrocardiogram characteristically shows variations in the TP interval only. When the heart is beating rapidly, as in exercise, the TP interval is shortened, the other intervals are little changed.

Premature contractions (extrasystoles) may arise from situations removed from the pacemaker (*ectopic foci*) at times when the muscle is no longer refractory and before the next excitatory process reaches them by the normal route. In such cases the position of the ectopic focus may be localized by inspection of the electrocardiogram. Thus, for example, if the ectopic focus is in the atria the P wave of the extrasystole may be abnormal in appearance but the ventricular complex is normal, since the process of activation spreads through the junctional tissue and ventricular muscle in the normal way (Fig. 28, 51). A premature contraction arising from an ectopic focus in a ventricle is identified by its grossly abnormal appearance in the electrocardiogram and by the absence of a preceding P wave since the electrical process does not begin in the atria (Fig. 28, 52).

Atrial flutter is the name given to the condition in which the atria contract at rates of some 180 to 350 per minute and the electrocardiogram (Fig. 28, 53) shows, therefore, a rapid and orderly succession of P waves which are, however, often somewhat unusual in form. The ventricles may respond only to every second or third of such impulses so that we recognize one ventricular complex for perhaps two or three P waves. Because the excitatory process, if accepted by the ventricles, proceeds by the normal route the QRS complexes are normal in appearance. Atrial fibrillation (Fig. 28, 54) is characterized by an absence of regularly formed P waves, since in this condition the atria are not contracting fully at any time, the muscle fibrillating at rates exceeding some 400 per minute. The

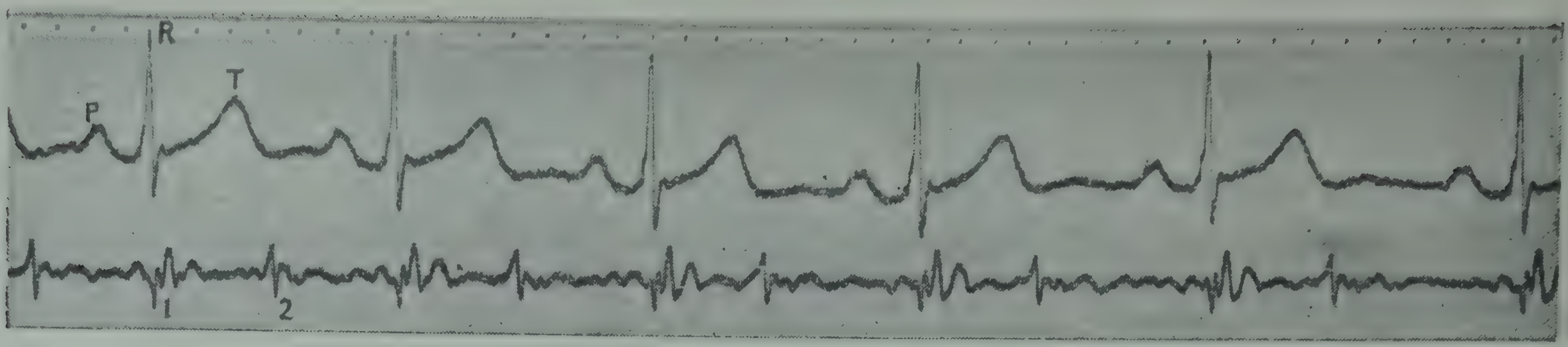


FIG. 28, 49

Electrocardiogram, lead II, and phonocardiogram lower tracing. The microphone was placed 1 inch medial to the apex beat: subject, male, 35: the dots along the top of the photograph are 0·1 sec apart.

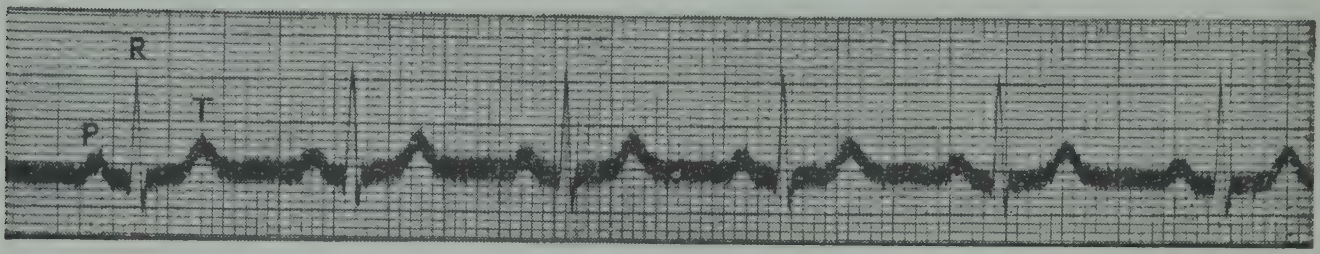


FIG. 28, 50

Normal electrocardiogram, lead II: 1 centimetre vertically = 1 mV. Time intervals in fifths and twenty-fifths of a second. (This and the following four diagrams have been reduced to about two-thirds of the original size.)

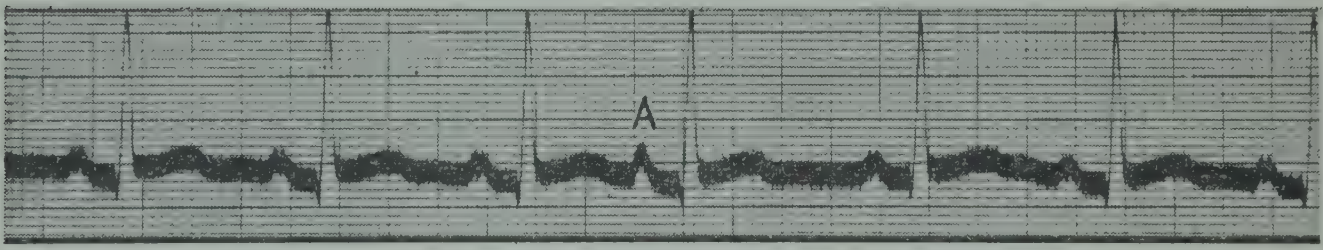


FIG. 28, 51

Electrocardiogram, lead II: showing an atrial extrasystole. The normal P waves with a slight notch in them follow one another at regular intervals. At the point marked A an atrial wave of abnormal shape occurs prematurely but is followed by a normal QRST complex.

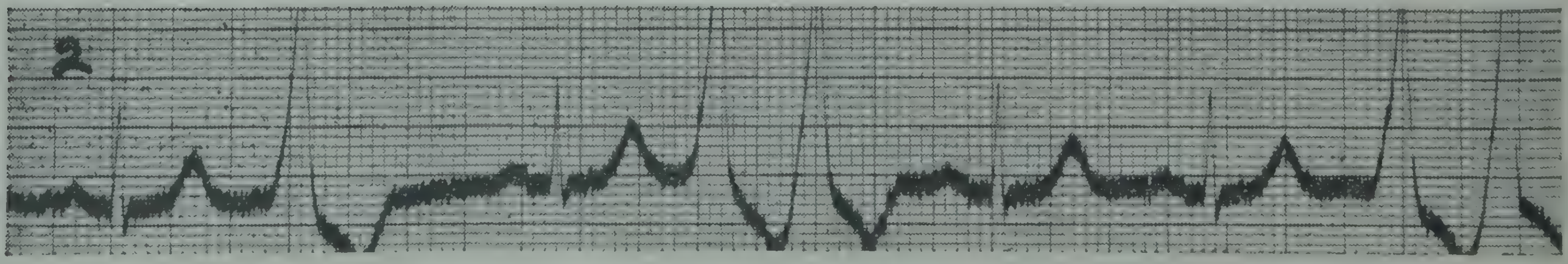


FIG. 28, 52

Electrocardiogram, lead II: showing ventricular extrasystoles. The bizarre ventricular complexes are not preceded by P waves.

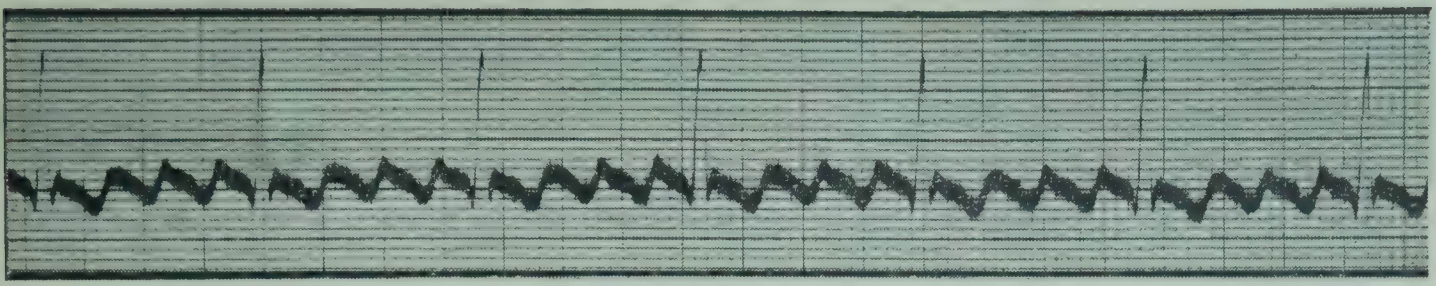


FIG. 28, 53

Electrocardiogram, lead II: showing atrial flutter. A regular succession of P waves at a frequency of 324 per minute indicate that the atrium is beating regularly at this rate but the ventricle responds only to each fourth beat. This is 4: 1 heart block. Time intervals in fifths and twenty-fifths of a second.

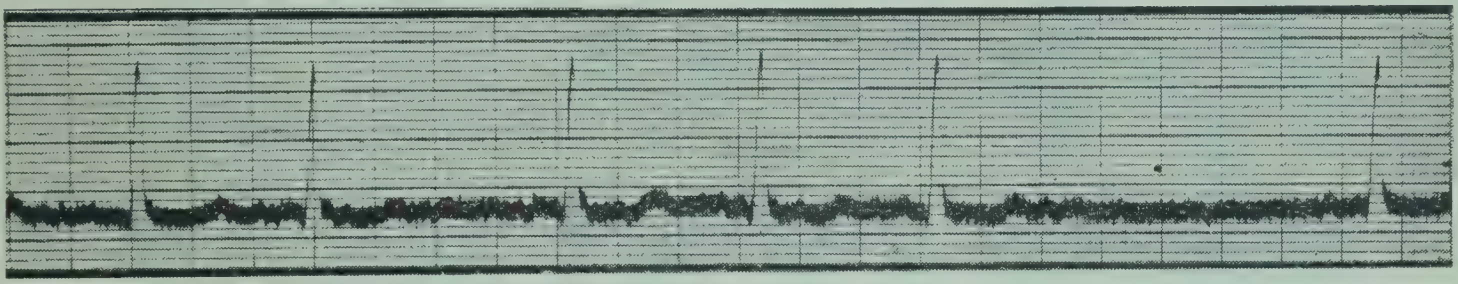


FIG. 28, 54

Electrocardiogram, lead II: showing atrial fibrillation. The base line is very irregular. No true P waves are present. The interval between QRST complexes is very variable. Time in fifths of a second.

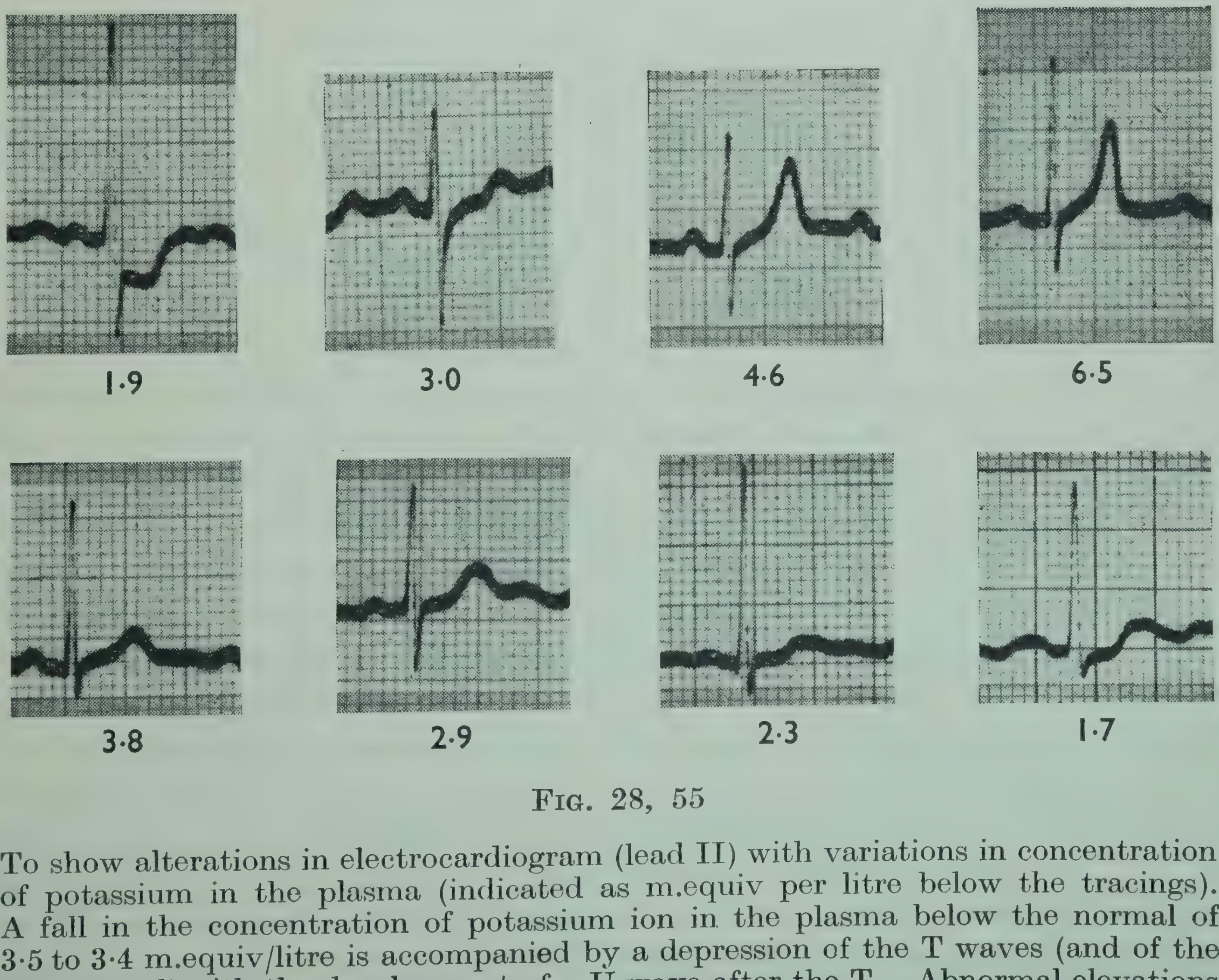


FIG. 28, 55

To show alterations in electrocardiogram (lead II) with variations in concentration of potassium in the plasma (indicated as m.equiv per litre below the tracings). A fall in the concentration of potassium ion in the plasma below the normal of 3·5 to 3·4 m.equiv/litre is accompanied by a depression of the T waves (and of the ST interval) with the development of a U wave after the T. Abnormal elevations of serum potassium are associated with large, peaked, upright T Waves.

ventricles respond irregularly to the rapidly repeated impulses reaching them from above so that the QRS complexes, though normal in form, are irregularly placed along the electrocardiogram.

Since the process of conduction of the electrical changes in the junctional tissue does not normally occupy more than 0·21 sec a PR interval above this value denotes impaired conduction. In complete heart block the atria and ventricles are completely dissociated from one another, the atria being activated from the pacemaker and the ventricles contracting to the slower rate of impulse formation in the A.V. node. In such a case the electrocardiogram shows regularly recurring P waves of normal form and, superimposed upon this, a regular succession of ventricular complexes at a slower rate (*idioventricular rhythm*). Since the activation process extends over the ventricles in the normal direction the QRS complexes are not deformed.

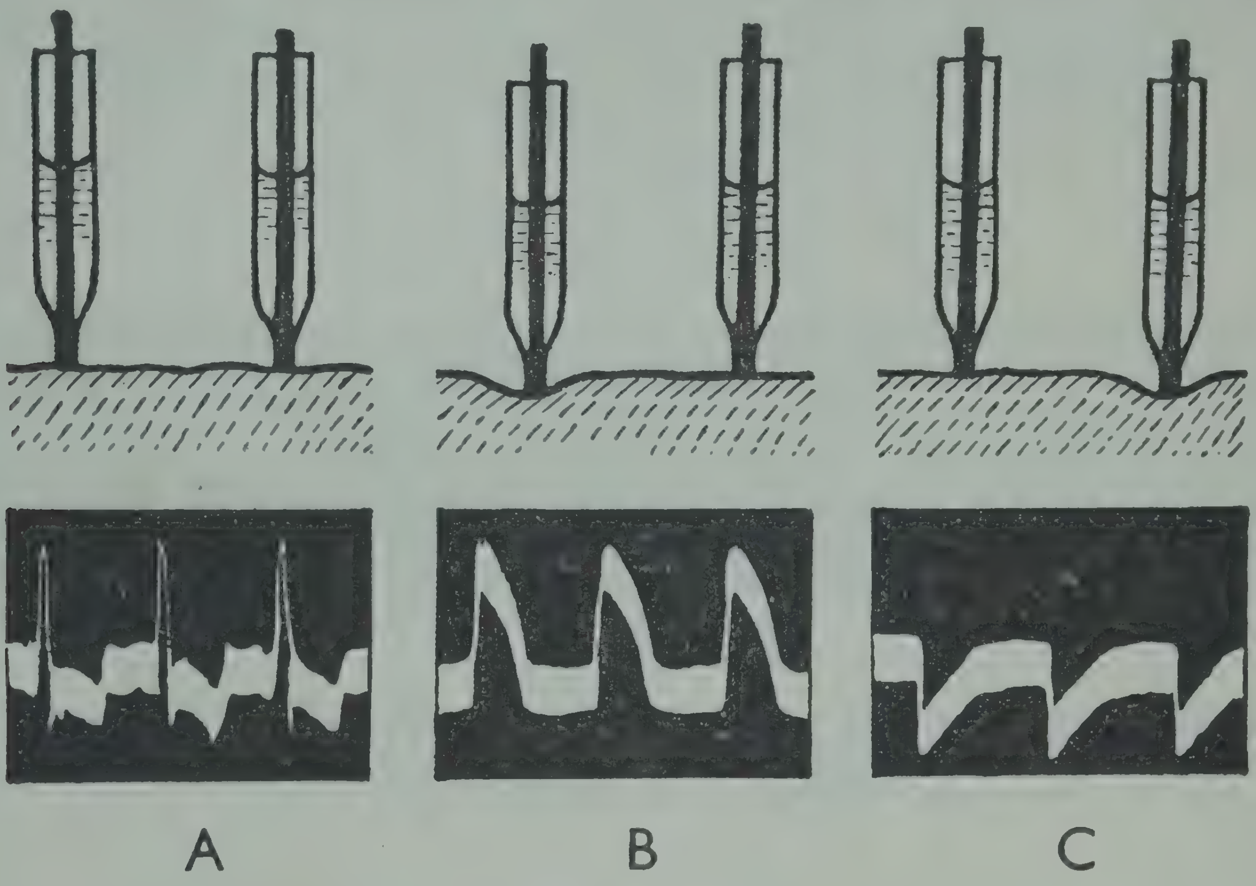


Fig. 28, 56

Three records showing the type of curve obtained from the surface of the dog's heart by two electrodes on its surface connected to a galvanometer (bipolar connexion). In segment A, both regions are uninjured; in segments B and C, injury has been produced in one or other of the regions beneath the electrodes by compression with electrode, as illustrated in the diagrams above the records. Segment A shows a diphasic (uninjured) electrogram and B and C show monophasic (injured) electrograms. The term electrogram is used to denote records from the heart's surface, in contradistinction to electrocardiograms obtained from regions distant from the heart's surface. (Reproduced from Katz (1946), *Electrocardiography*, 2nd ed. London: Henry Kimpton.)

Nowadays, perhaps the most important use of the electrocardiograph is as an aid to the diagnosis of disease of the coronary arteries which may reduce the blood supply to the muscle or conducting tissue of the heart by leading to narrowing of the vessels or their occlusion by blood clot. The interpretation of the findings in coronary disease is largely empirical, but the abnormalities in the first few days following occlusion seem to depend on the fact that if a tissue is injured the polarization of the cell membranes cannot be maintained and the galvanometer records a steady difference of potential as if the condition pictured in Fig. 28, 46, *b*, persisted. When the injured part of the tissue is activated a monophasic wave and not a biphasic wave is recorded (Fig. 28, 56). Thus in coronary disease the ST segment may be elevated or depressed from its normal position and T is often inverted. The position in the heart of the damaged area of muscle tissue determines the shape of the electric field and, if the alterations in the field affect the electrodes, such areas of damage can often be quite accurately localized by an exploring electrode moved over the præcordium. In so-called silent areas, however, damage causes no alteration in the electrocardiogram. As the tissue heals or dies the abnormalities described above disappear but the electrocardiogram never becomes completely normal again because of the presence in the muscle of tissue that cannot be activated.

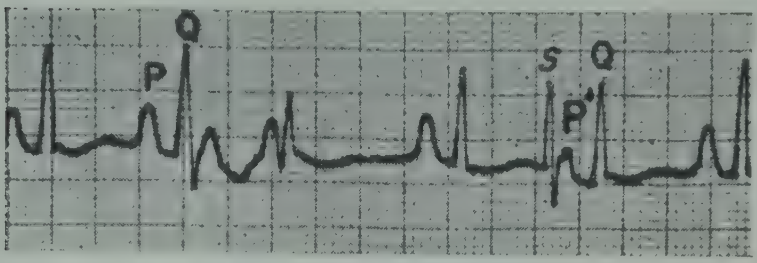
ATRIAL FIBRILLATION AND FLUTTER

At atrial systole in a normally functioning heart all the atrial muscle fibres contract practically simultaneously and, in the ventricular systole which follows after a short interval, all the ventricular muscle fibres contract at almost the same instant. In certain diseases of the heart, or after strong electrical stimulation of an animal's heart, this orderly contraction may be replaced by very irregular activity, in which the individual muscle fibres contract in a disorderly sequence all out of step with one another so that the surface of the heart shows a fine rapid tremulous movement (*fibrillation*). Since the affected chambers of the heart do not contract as a whole their pumping action is lost. If the ventricles fibrillate the blood is not pumped out into the systemic or pulmonary circulations and death results from anoxia of the vital centres. If, however, fibrillation is confined to the atria the ventricles still force the blood round the circulation and a patient with atrial fibrillation may have a reasonably efficient circulation. Mackenzie found that the *a* waves of the jugular pulse were absent in such cases and supposed that the atrium was paralysed, but later electrocardiographic investigations showed that this is not so.

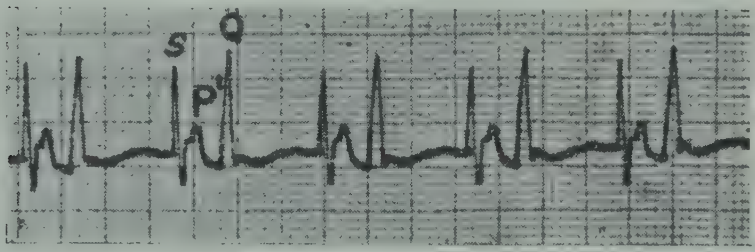
If the atrium of the exposed heart of the dog is subjected to a brief period of rapid electrical stimulation (300 to 600 per minute) the atria may change their rhythm and beat about 300 times per minute (*atrial flutter*). If the spread of the excitation wave is followed by electrocardiographic methods it is found to circulate continuously along a path around the mouths of the great veins and Lewis believed that this so-called circus movement was the basis of both flutter and fibrillation.

The mechanism of atrial flutter and fibrillation has been studied by Prinzmetal and his colleagues with a high-speed cinematograph (2000 frames per second) and a cathode-ray oscillograph which is capable of recording electrical changes occurring as rapidly as 20,000 per minute. In atrial fibrillation there are continuously present minute irregular contractions (M) and also larger rhythmic contractions (L). The M contractions occur throughout the atrial musculature, each one involving an area of the wall from 0·03 to 3 mm in diameter. Superimposed on these are the L contractions which sweep across the atrium fairly regularly at rates of 400 to 600 per minute. They do not follow a circus pathway and no daughter waves are seen; it is these L contractions which give the appearance of fibrillation to the unaided eye. Cutting or burning of the atrium which would interrupt Lewis's circus movement does not abolish the fibrillation. Atrial fibrillation is, therefore, due to the discharge of impulses

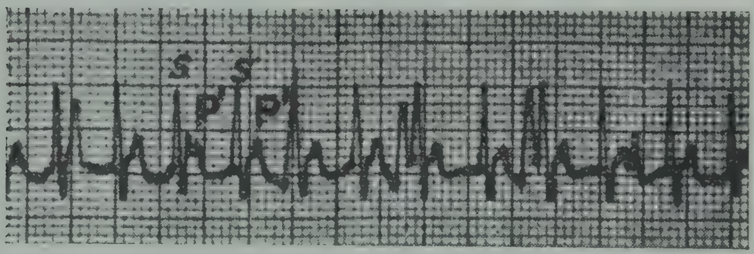
Normal sinus rhythm with atrial extrasystoles



Atrial tachycardia (atrial rate = 176/min)



Atrial flutter (atrial rate = 428/min)



Atrial fibrillation

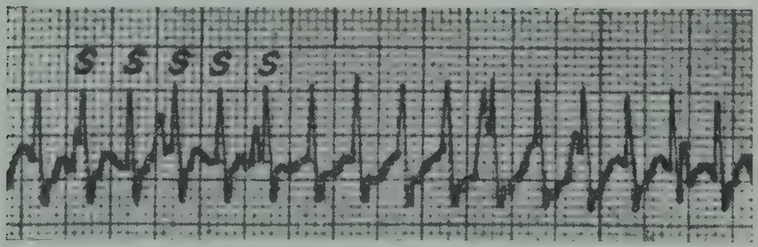


Fig. 28, 57

Electrocardiograms showing gradual progression from atrial extrasystoles (S) to fibrillation. Ordinate: 1 cm≡1mV. Abscissa marked in fifths and twenty-fifths of a second. (M. Prinzmetal (1952), *Canad. med. J.* **66,** 535.)

at rates of over 400 per minute from a focus in the atrial muscle which, since it is other than the S.A. node, is said to be *ectopic*.

Prinzmetal has similarly failed to find evidence of circus movement in atrial flutter, and interruption of a hypothetical circus pathway by burning or cutting did not arrest the disturbance. Prinzmetal believes that all of the atrial arrhythmias arise in a single ectopic focus. At rates of impulse formation in the ectopic focus below the normal sinus rate premature atrial systoles are produced; when the rate of impulse formation increases to 100 to 250 per minute then atrial tachycardia is the result. Flutter and fibrillation are found at rates of impulse formation of approximately 300 and 400 per minute respectively (Fig. 28, 57).

In both atrial flutter and fibrillation the rate of stimulation is so rapid that the ventricles are incapable of responding to every impulse reaching them from the atria. When flutter is present the ventricles usually respond in a regular manner to every third or fourth atrial contraction (Fig. 28, 53) while in fibrillation the response of the ventricles is irregular (Fig. 28, 54) and often rapid.

The alkaloid quinidine is often effective in arresting atrial arrhythmia of whatever type. This is probably because the drug depresses the excitability and conductivity of cardiac muscle and so leads to a slower rate of impulse formation.

REFERENCES

Altschule, M. D. (1949). *Physiology in Disease of the Heart and Lungs.* Harvard University Press.

Anrep, G. B. (1936). *Studies in Cardio-Vascular Regulation.* Stanford University Press.

Bing, R. J., Hammond, M. M., Handelsman, J. C., Powers, S. R., Spencer, F. C., Eckenhoff, J. E., Goodale, W. T., Hafkenschiel, J. H. & Katy, S. S. (1949). Measurement of coronary blood flow, oxygen consumption and efficiency of the left ventricle in man. *Amer. Heart J.* **38,** 1–24.

Clark, A. J. (1927). *Comparative physiology of the heart.* London: Macmillan.

Cruickshank, E. W. H. (1936). Cardiac metabolism. *Physiol. Rev.* **16,** 597–639.

Dawes, G. S. (1951). Reflexes from the heart and lungs. *Brit. med. Bull.* **8,** 324–330.

Gregg, D. E. (1950). *Coronary Circulation in Health and Disease.* London: Kimpton.

Harrison, T. R. (1935). *Failure of the Circulation.* Baltimore: Williams Wilkins Co.

Heymans, C. (1950). *An Introduction to the Regulation of Blood Pressure and Heart Rate.* Oxford: Blackwell.

Katz, L. N. (1946). *Electrocardiography.* 2nd ed. London: Henry Kimpton.

Lewis, T. (1925). *Mechanism and Graphic Registration of the Heart Beat.* 3rd ed. London: Shaw & Sons.

Lewis, T. (1949). *Electrocardiography and Clinical Disorders of the Heart Beat.* London: Shaw & Sons.

McMichael, J. (1952). The dynamics of heart failure. *Brit. med. J.* **2,** 525–529 and 578–582.

Starling, E. H. (1918). The law of the heart. *Linacre Lecture,* 87.

Symposium on the regulation of the performance of the heart (1955). *Physiol. Rev.* **35,** 91–168.

Wiggers, C. J. (1952). *Circulatory Dynamics; A Physiological Study.* New York: Grune & Stratton.

CHAPTER 29

DYNAMICS OF THE PERIPHERAL CIRCULATION

STRUCTURE OF THE BLOOD VESSELS

ALL the blood vessels of whatever size have a smooth inner lining of flattened *endothelial cells* joined edge to edge. This inner lining is continuous from the arteries through the capillaries to the veins and to the internal lining of the heart (the *endocardium*). The capillary walls consist of this endothelial layer and very little else (Fig. 29, 1).

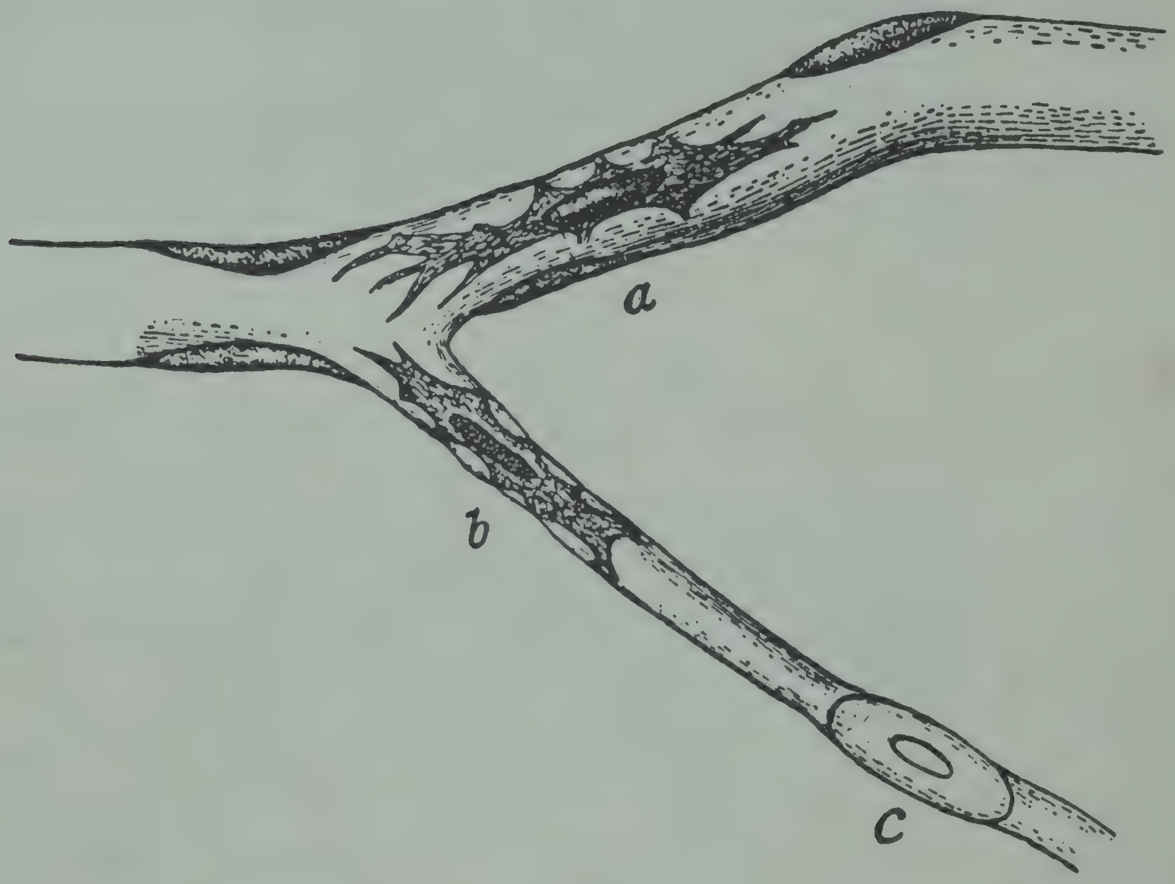


FIG. 29, 1

Structure of capillaries

a and *b* are two branched Rouget cells, as seen in the capillaries of a living newt larva. *c* is a nucleated red blood corpuscle. The other cells are endothelial cells. (*After* B. Vimtrup (1922), *Z. ges Anat.* **65,** 150.)

The arteries have three main coats and the following description refers mainly to a medium-sized artery (Figs. 29, 2 and 3). The endothelial layer is surrounded by an elastic layer, the two together forming the *tunica intima*, or inner coat. After death, when the internal pressure falls, the elastic tissue contracts and forms the wavy inner border of an artery seen in histological preparations. The tunica intima is surrounded by smooth muscle fibres which pass circularly round the vessel, forming the *tunica media*, or middle coat. Finally there is an external coat, or *tunica adventitia*, of fibrous tissue which blends with the general fibrous tissue of the body. In the larger arteries the elastic layer takes the form of a fenestrated membrane and there is also a certain amount of subendothelial fibrous

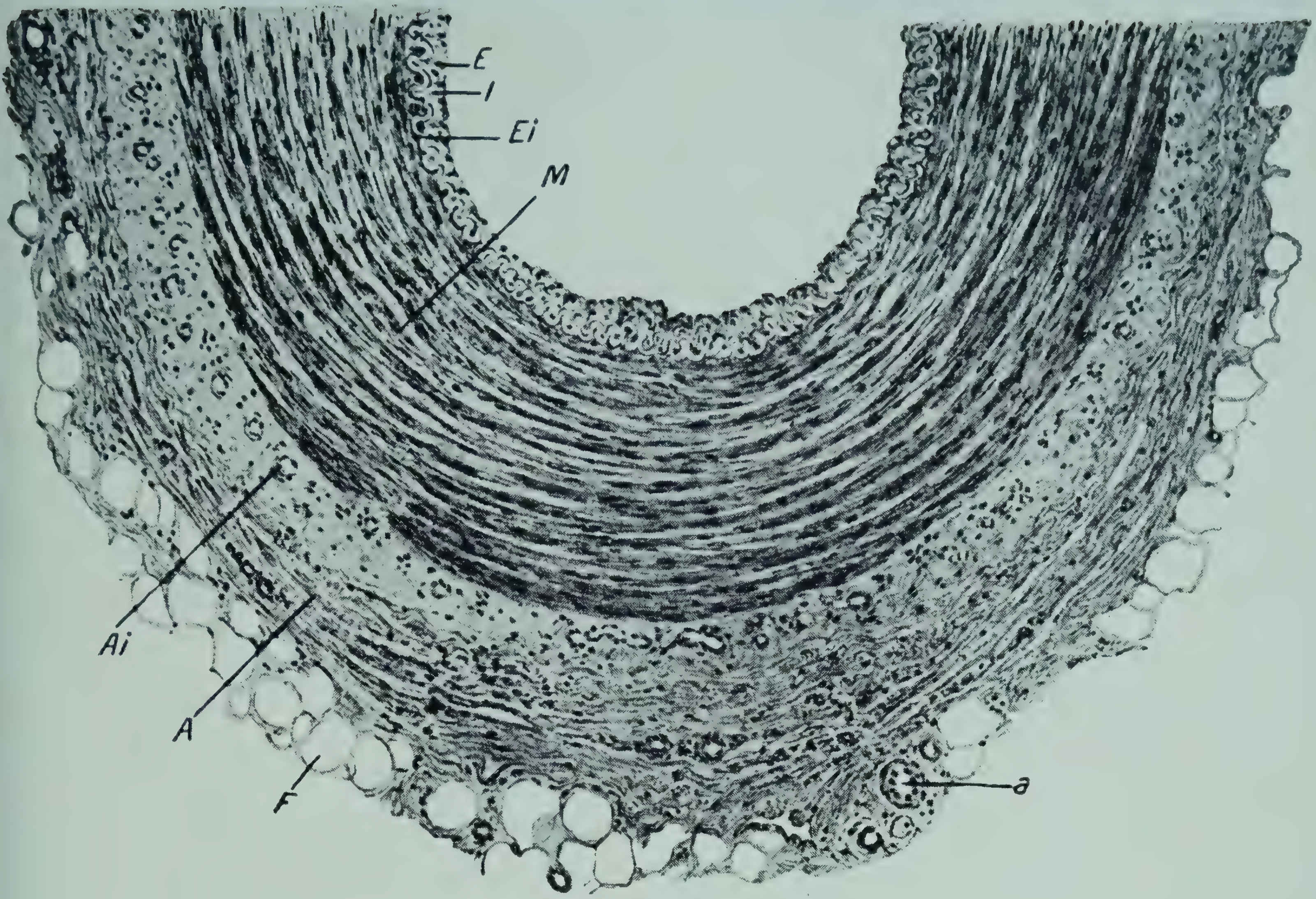


FIG. 29, 2

Half of a cross-section of the volar digital artery of an executed man: *A*, Adventitia; *Ai*, internal layer of the adventitia with many vasa vasorum; *E*, endothelium; *Ei*, internal elastic lamina; *I*, intima; *M*, muscular coat (media); *F*, fat tissue; *a*, cross-section of a small artery. 80 ×. (*After* Schaffer.)

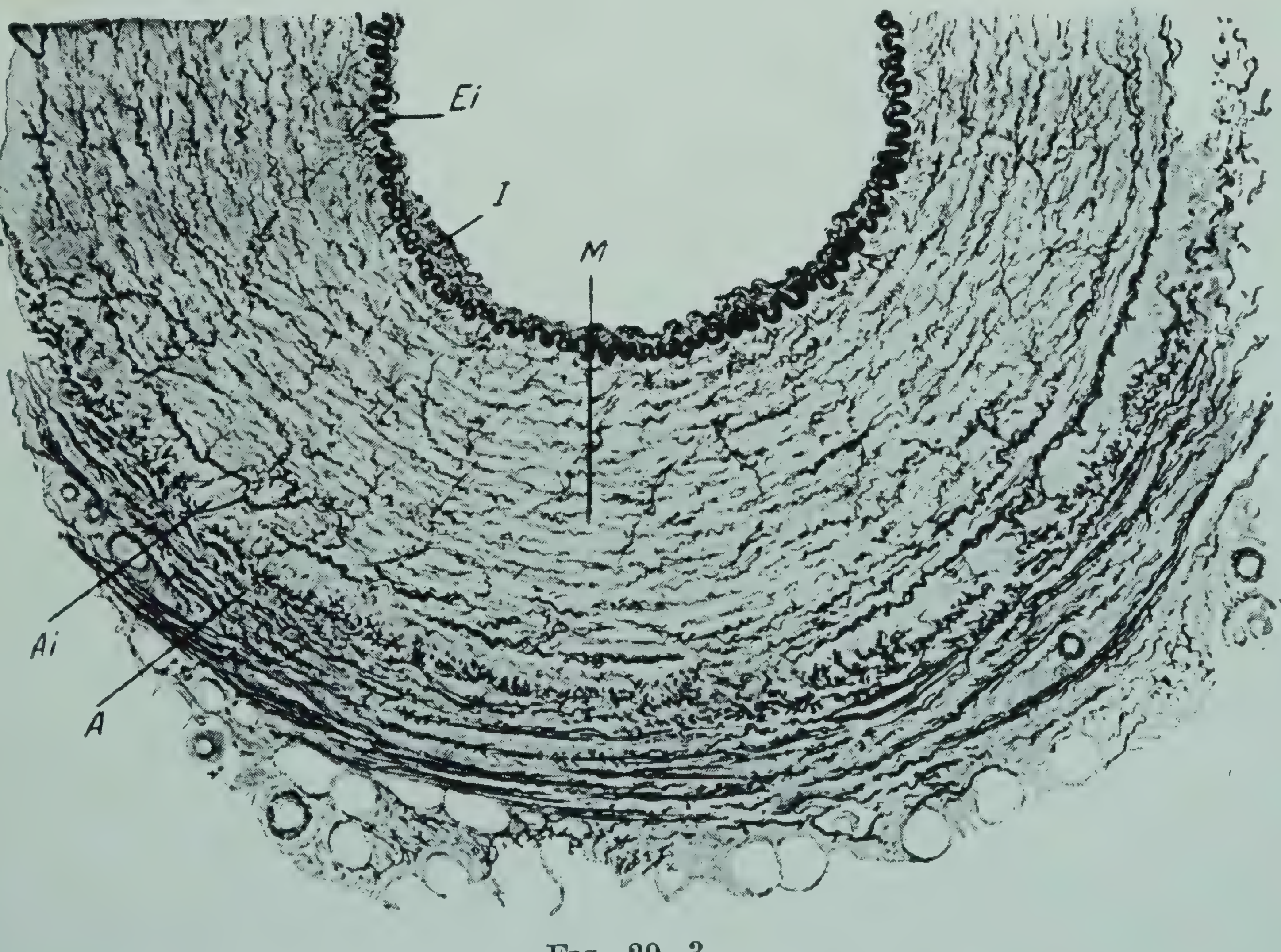


FIG. 29, 3

An analogous cross-section as in Fig. 28, 2, stained with acid orcein to show the elastic tissue. Abbreviations and magnification as in the preceding figure. (*After* Schaffer.)

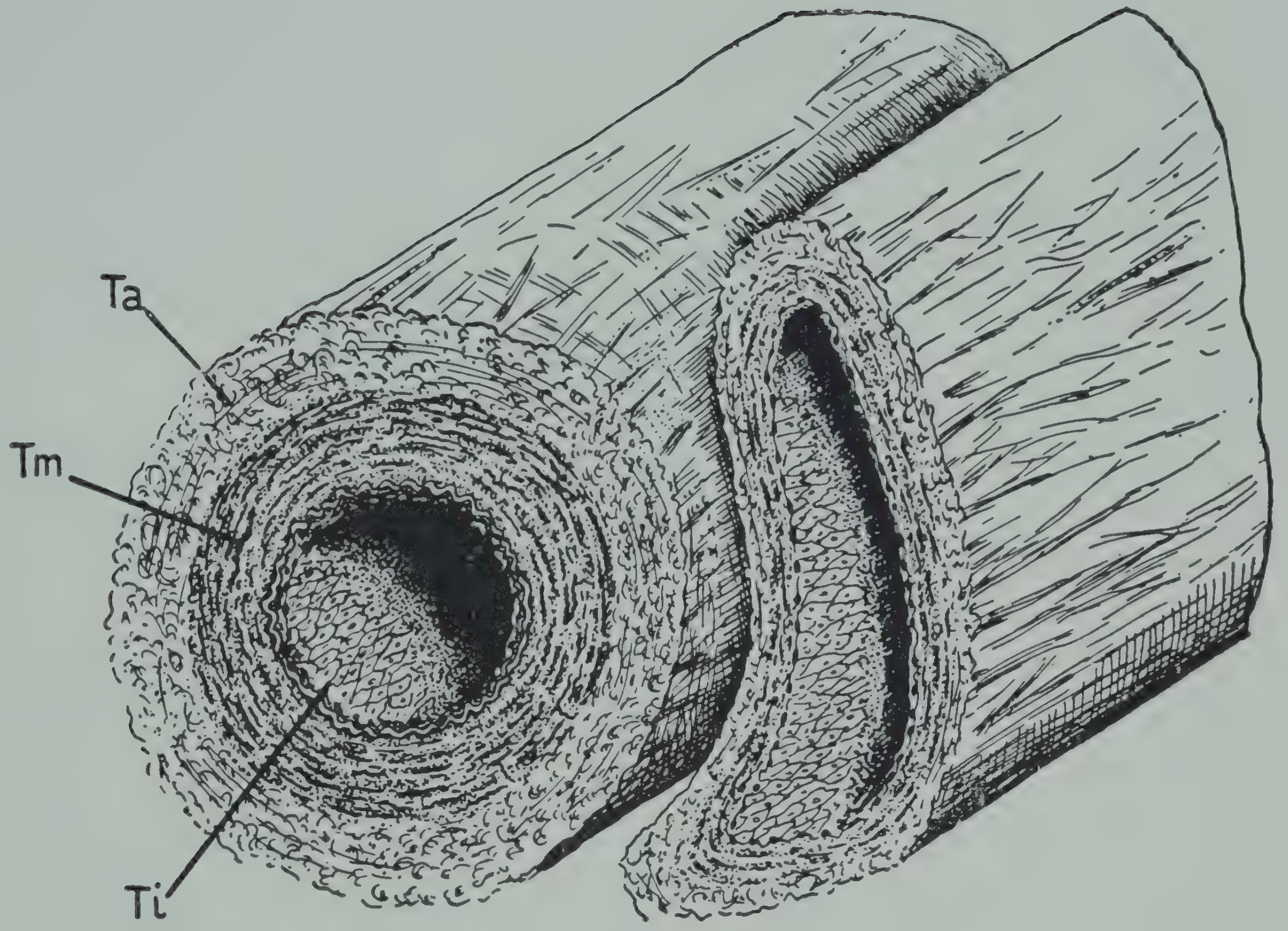


Fig. 29, 4

A diagram constructed to emphasize schematically the contrast between the cross-sectional appearance of an artery and its corresponding vein. *Ta*, tunica adventitia. *Tm*, tunica media. *Ti*, tunica intima (W. E. Le Gros Clark (1945), *The Tissues of the Body*, p. 173. Oxford: Clarendon Press).

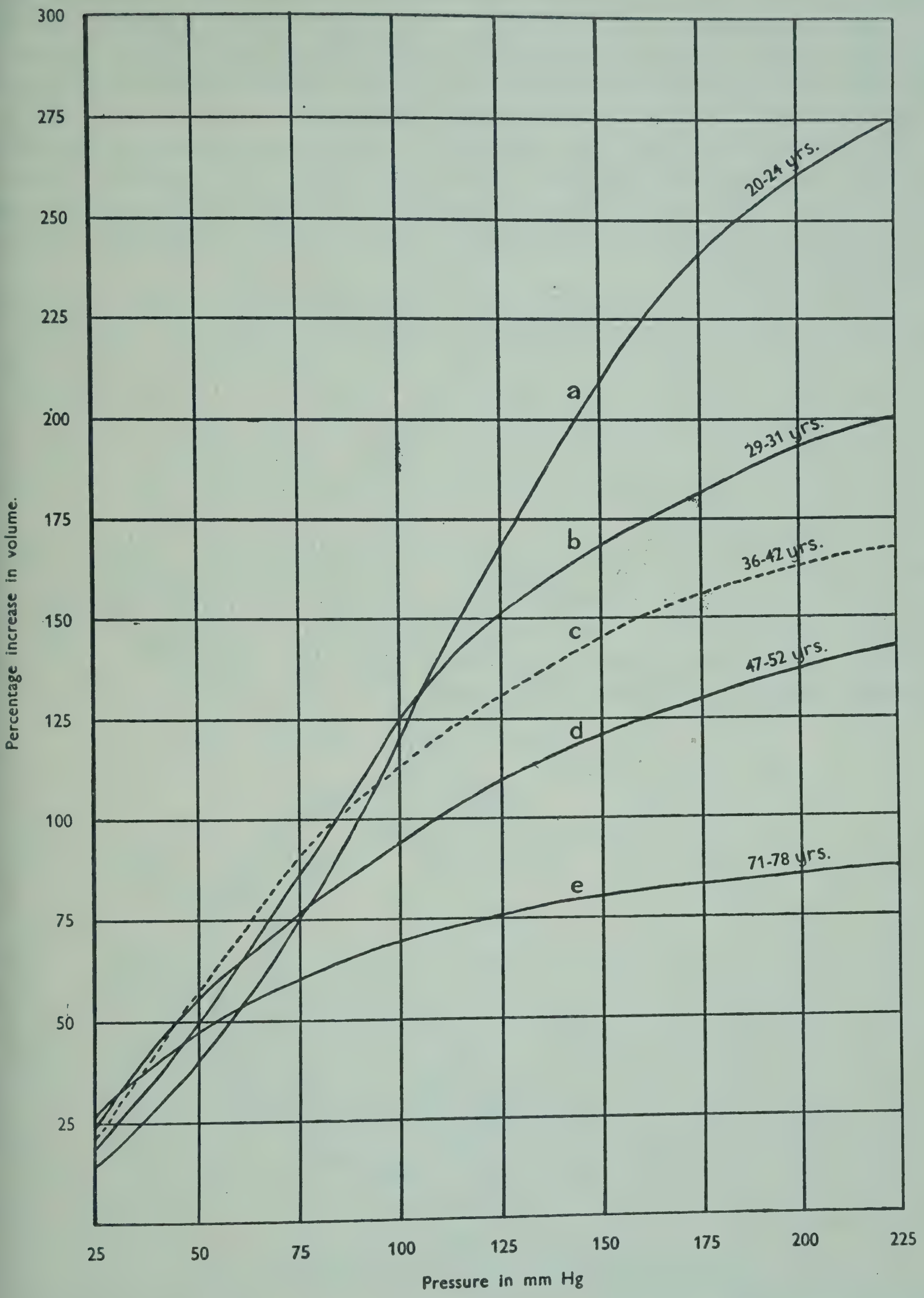


FIG. 29, 5

Mean volume-elasticity curves of thoracic aortas in various age groups (P. Hallock & I. C. Benson (1937), Elastic Properties of Human Aorta, *J. clin. Invest.* **16,** 597).

tissue. The muscle fibres of the middle coat are reinforced by a network of elastic fibres and the outer coat contains elastic fibres in addition to white fibrous tissue (Fig. 29, 3). The large arteries, and particularly the aorta, thus contain a relatively greater proportion of elastic tissue in their walls than do the smaller arteries. The walls of the smallest arteries of all, the arterioles, are almost entirely muscular.

The walls of the veins are very much thinner than those of the corresponding arteries although their internal diameter is generally greater (Fig. 29, 4). The elastic tissue is usually quite inconspicuous and some veins, for example those of the brain, bones and placenta, are said to have no muscular tissue.

At the point where an artery gives off a branch, or divides into two, the combined cross-sectional area of the two divisions is somewhat greater than that of the original artery. The total cross-sectional area of the vascular system, therefore, gets gradually larger as the vessels themselves become smaller so that at the capillaries it may be one thousand times that at the aorta. It is probable, however, that under normal conditions a large number of the smaller vessels are closed, so that the total effective cross-section is reduced.

The veins collapse when the internal pressure is zero and are easily distended by a small rise of pressure; the arteries, on the other hand, remain patent when the pressure is zero and they withstand a considerable internal pressure with a relatively small increase in volume. This is indicated in Fig. 29, 5 which also shows that the elasticity of the aorta gradually diminishes as the age increases.

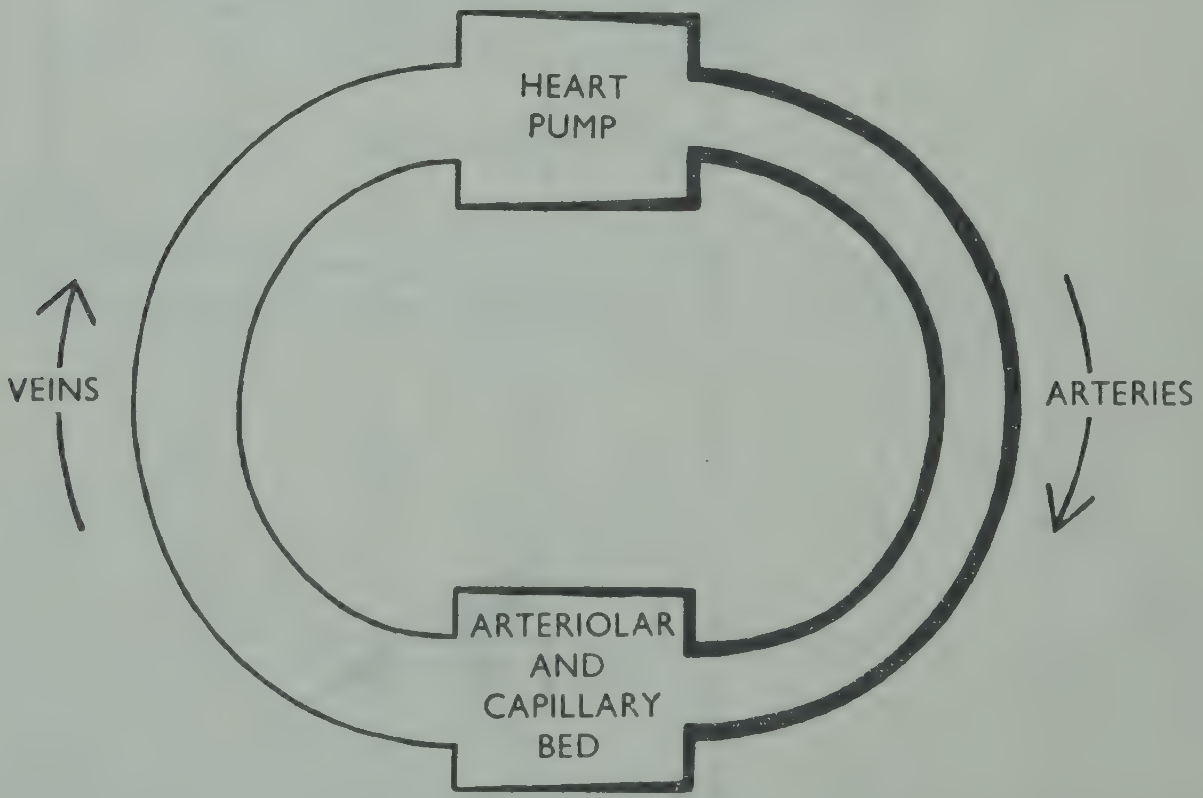


Fig. 29, 6

Diagram to illustrate the capacity of the vascular system.

The Capacity of the Vascular System

The effect of alterations of capacity of the vascular system can be illustrated by referring to Fig. 29, 6. At any given rate of pumping and at a fixed output per beat a steady state is soon reached in which the heart delivers to the arteries as much blood as reaches it from the veins. If extra fluid is injected into the veins there is very little rise of venous pressure because the veins are thin-walled and easily distended. If now the heart pumps more quickly, so that the venous pressure is brought back to its initial value, the arterial pressure rises considerably because an increase of arterial capacity can be produced only by a great increase of pressure (Fig. 29, 5). In the living body, however, additional fluid might be accommodated by a dilatation of the capillary bed with very little alteration in either venous or arterial pressure. Suppose now, in the model, that blood is withdrawn from the venous side. This results in a small fall of venous pressure but a large fall of arterial pressure. In the intact animal, however, although the venous pressure falls, the fall in arterial pressure is prevented by arteriolar constriction which diminishes the capacity of the vascular system; in other words the vascular system is able to alter its capacity to accommodate its contents without gross changes in hydrostatic pressure.

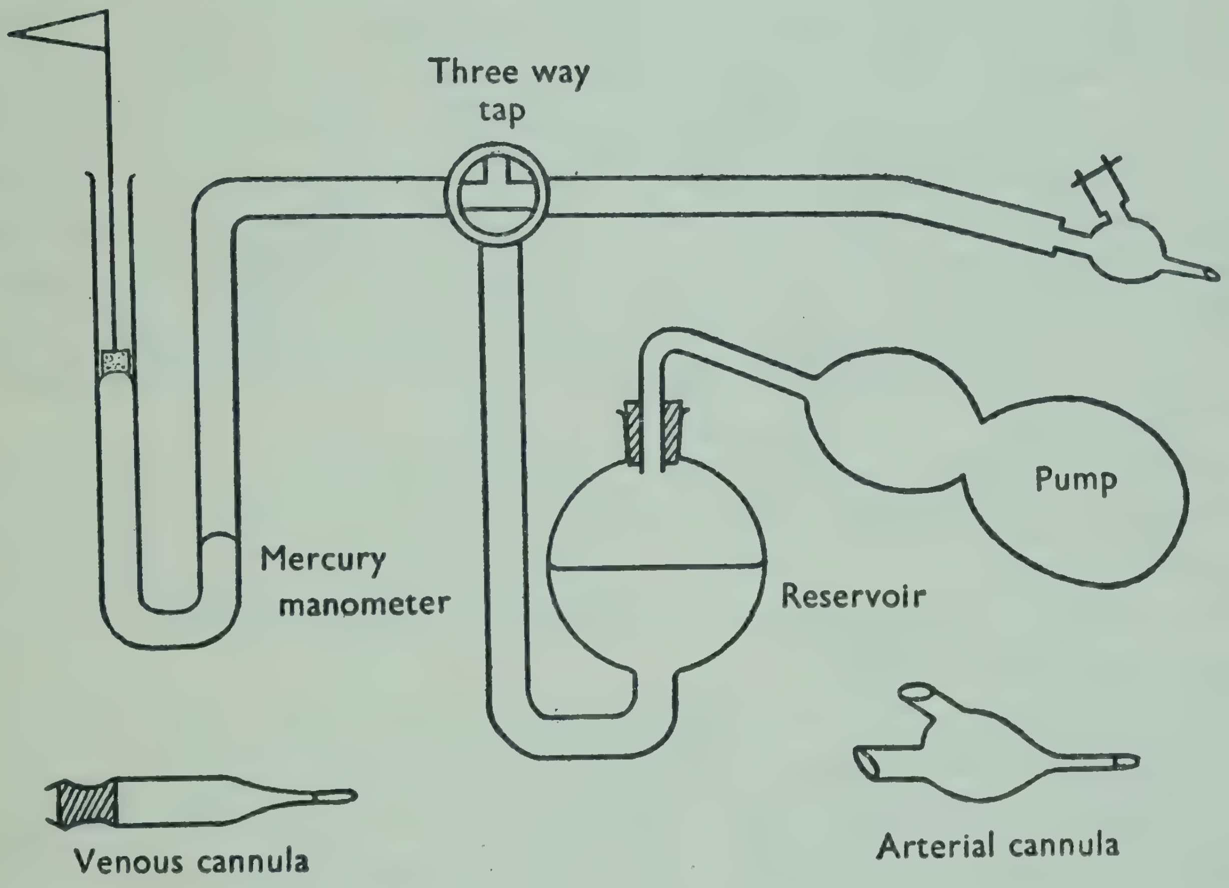


Fig. 29, 7

Apparatus for recording blood pressure in animals. The reservoir and the tubes connected to it by the three-way tap are filled with anticoagulant fluid. A clamp is placed on the rubber tube near the arterial cannula, the three-way tap is rotated to connect all three tubes (**T**-position) and the pressure is raised by the pump to, say, 150 mm Hg. The three-way tap is then turned to the position ⊥ shown in the diagram. The cannula is tied in the artery. When the clamp is released the blood in the artery is connected via a column of anticoagulant fluid to the mercury manometer (G. H. Bell (1952), *Experimental Physiology*, 5th ed. p. 217. Glasgow: Smith & Son).

PRESSURE IN THE VESSELS

Arterial Pressure.—It is well known that the blood escapes from a cut artery under considerable pressure, and the first attempt to measure the pressure was made by Stephen Hales in 1732. After occluding the artery with a temporary ligature he tied a brass pipe into the femoral artery of a mare and connected to it a glass tube nine feet long. When he loosened the ligature on the artery the blood rose 8 ft. 3 in. (about 250 cm) above the level of the heart. He noted that alterations in pressure occurred with each beat of the heart and with the movements of respiration. Although these results are of great interest the method is not very useful because the blood clots in the manometer in a few minutes.

Nowadays for such an experiment the artery is connected by a tube containing an anticoagulant, for example saturated sodium carbonate or 3 per cent. sodium citrate, to a mercury manometer as in the diagram (Fig. 29, 7). The difference in level of the two mercury surfaces in the U-tube gives the pressure which is usually expressed in mm Hg. Owing to the large inertia of the column of mercury the oscillations with each beat of the heart are small. The instrument, therefore, records not the maximum and minimum pressures within the artery but the *mean arterial pressure* which is the average pressure within the artery with respect to time. When arterial pressure is measured with a more rapidly responding manometer, such as that described in discussing ventricular pressure (p. 435), the oscillations are very much greater. The maximum or systolic pressure is reached very rapidly after ventricular systole; thereafter the pressure falls relatively slowly to its minimum or diastolic value (Figs. 28, 28; 28, 29). Because of the quick rise and slow fall of the pulse wave the average of the systolic and the diastolic pressures obtained with a rapidly responding membrane manometer is not the same as the mean arterial pressure. The mean arterial pressure is in fact closer to the diastolic than to the systolic pressure, approximating to the diastolic pressure plus one third of the pulse pressure.

The arterial pressure of man was first measured by such a method in 1856 by Faivre. Nowadays very accurate and continuous records of the human blood pressure can be obtained by condenser, transducer or strain-gauge manometers communicating through a small needle with the inside of an artery (see Fig. 29, 20).

Arterial Pressure in Man.—The direct methods just described are obviously unsuitable for routine clinical use but the blood pressure can be estimated in man indirectly by a development of the method originally invented by Riva-Rocci in 1896. The sphygmomanometer in general use nowadays consists of a mercury manometer with one wide and one narrow limb (Fig. 29, 8). The pressure to be measured is applied to the wide limb; the mercury in it moves down so little that the calibrated divisions on the narrower limb are nearly twice as far apart as in the original type of manometer which had limbs of equal cross-section. A rubber bag of standard size (18 cm × 12 cm for adults) covered with cloth is wrapped round the upper arm, leaving the cubital fossa exposed, and connected as shown in the figure to the manometer and pump with which it is rapidly inflated to a point above the systolic blood pressure. The valve of the air leak (V) is then unscrewed a little to allow the pressure in the bag to fall gradually. The arterial pressure can then be assessed in one of three ways. The first appearance of pulsation in the arteries distal to the cuff may be detected with the fingers, the pressure in the cuff at that time being some 10 to 15 mm Hg below the true systolic blood pressure. Secondly the pulsations of the air in the cuff may be recorded by some kind of oscillometer; then the first appearance of oscillations indicates the systolic blood pressure and the point of maximum oscillation the diastolic blood pressure.

More usually the blood pressure is measured by the *auscultatory method*. The bell of a stethoscope is placed lightly over the brachial artery at the bend of the elbow. The pressure in the cuff is raised rapidly until the radial pulse disappears. When the pressure is allowed to fall a series of sounds is heard, first described by Korotkov in 1905 (Fig. 29, 9). If the pressure in the bag is above the systolic pressure then the artery is occluded throughout the pulse cycle. When the pressure falls a little below the systolic level (line A) the artery opens momentarily when the internal pressure is greater than the external (cuff) pressure and a blowing sound is heard at each pulse. As the pressure is further reduced the blowing sound is gradually succeeded by a distinct tapping which becomes louder as the pressure is reduced. As the pressure in the cuff is allowed to fall, preferably about 2 or 3 mm Hg per second, the character of the sound suddenly changes from a sharp tapping to a softer blowing noise (Fig. 29, 9, line C). Finally the sounds disappear, usually quite suddenly, when the pressure in the bag falls below diastolic blood pressure (line D) because below this pressure the internal pressure in the brachial artery is always greater than the pressure applied externally by the bag; the flow of blood along the artery is not interrupted and usually no sound is heard. The point at which regularly recurring sounds begin is taken as the systolic blood pressure (S.B.P.). The point at which the sound disappears completely is taken as the diastolic blood pressure (D.B.P.). Readings obtained by the auscultatory method were compared by Steele with readings obtained directly in the same person by means of a membrane manometer connected to a hollow needle in the lumen of the artery. He found that the point at which the sounds disappeared completely gave the best estimate of D.B.P.; it has, however, been customary in Britain to take the reading of diastolic

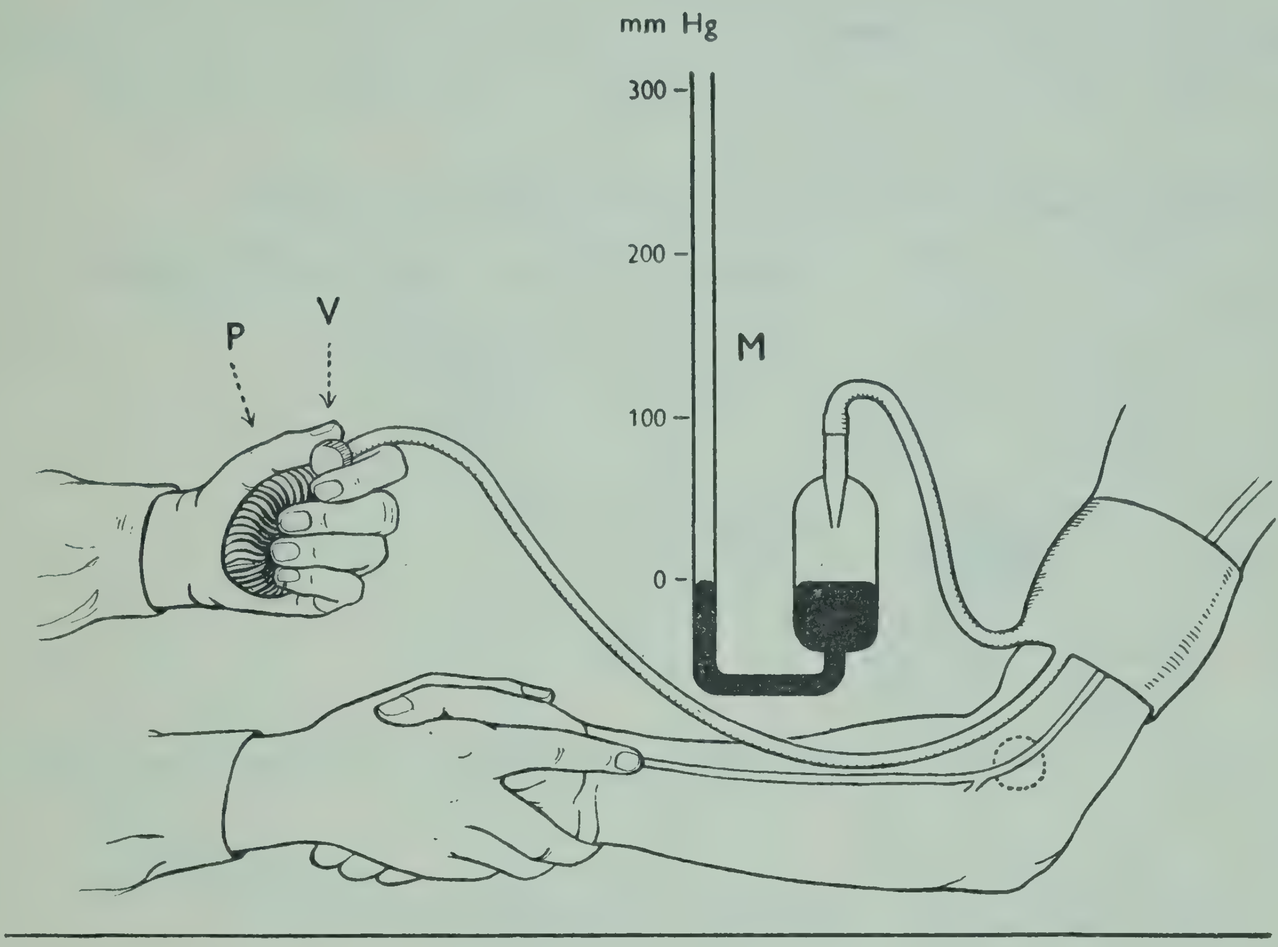


FIG. 29, 8

Method of measuring blood pressure in man. P = pump. V = screw for releasing air pressure. M = manometer. The dotted circle in the cubital fossa indicates the position of the bell of the stethoscope in the auscultatory method.

blood pressure when the sounds suddenly alter in character (Fig. 29, 9, line C). The difference between S.B.P. and D.B.P. is called the *pulse pressure*.

When the subject is lying the arterial pressures in the brachial and femoral arteries are approximately the same. With the subject standing, however, the femoral pressure is higher than that in the brachial artery, the difference being due to the pressure of the column of blood equal in height to the vertical distance between one artery and the other. The effect of posture on the blood pressure is shown in Table 29, 10.

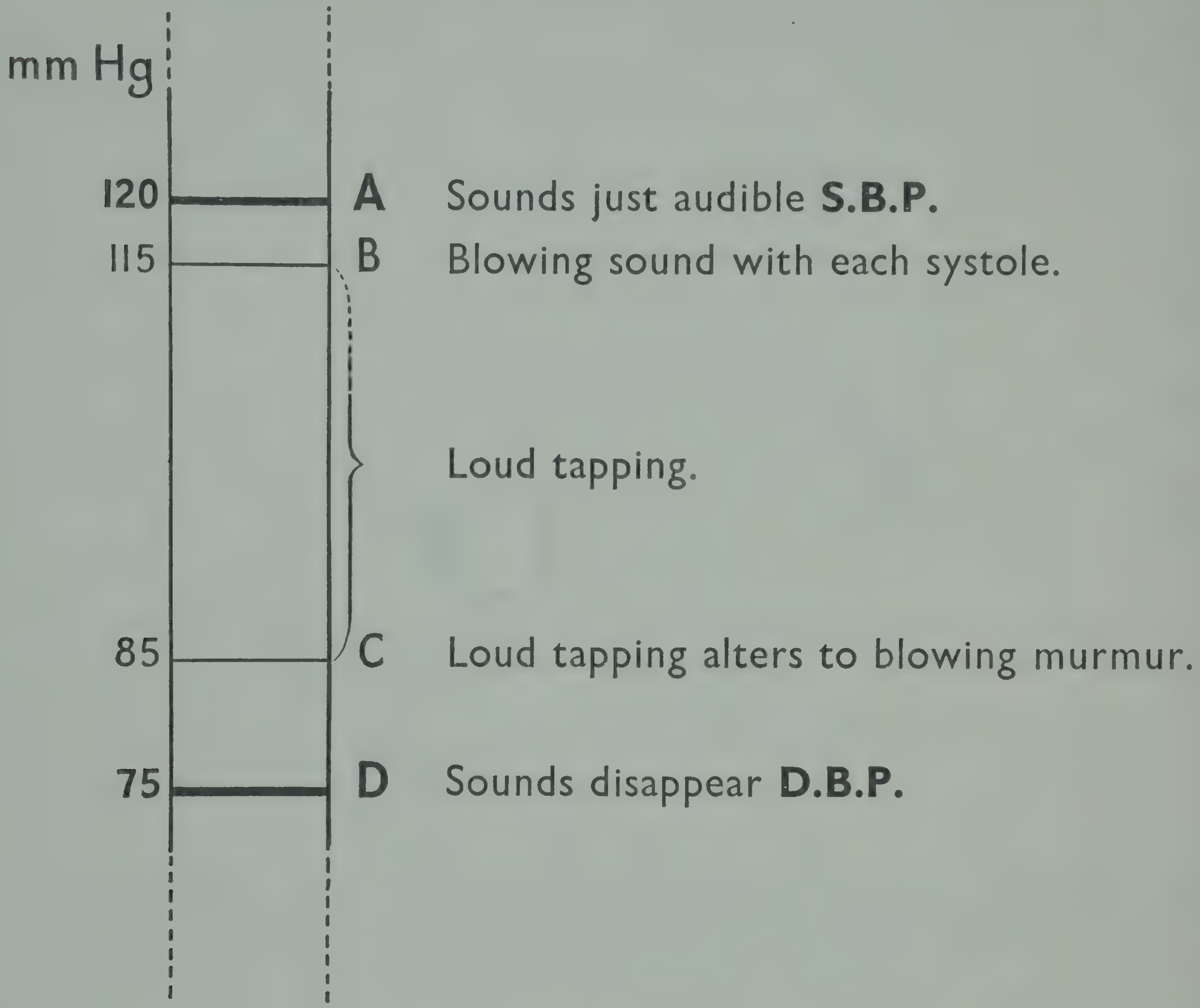

mm Hg
120
115
85
75
A Sounds just audible **S.B.P.**
B Blowing sound with each systole.
Loud tapping.
C Loud tapping alters to blowing murmur.
D Sounds disappear **D.B.P.**

Fig. 29, 9

The auscultatory method of determining arterial blood pressure. Diagram explained in the text.

In any individual arterial pressure is not constant but subject to appreciable variations over short intervals of time (Table 29, 12). The pressure recorded under ordinary conditions of life is referred to as the *casual* blood pressure. The figure is higher, sometimes much higher, than the *basal* blood pressure, that is the reading obtained ten to twelve hours after the last meal of the previous night, and after resting half an hour in a warmed room.

Even in the basal state there are fluctuations of the arterial pressure with respiration. These are readily observed in a manometer communicating with the large artery of an animal and were first noted by Stephen Hales in the experiment referred to on p. 483. Continuous records of arterial blood pressure made in man with electrical manometers show similar features (Fig. 29, 11; 31, 2) which are believed to be due to variations in both stroke volume and peripheral resistance. The precise relationship between arterial blood pressure and respiration depends on the respiratory rate; at normal rates of breathing the blood pressure falls during most of inspiration but when breathing is slow inspiration is accompanied by a rising blood pressure.

Many attempts have been made to find values between which a subject's blood pressure could be regarded as normal, but all such efforts have been, for various reasons, unsatisfactory. Arterial pressure increases with age, more in some subjects than in others, and no dividing line can be set at any age below which the pressure is normal and above which it is abnormal.

TABLE 29, 10

Subject	Posture	Brachial Artery Pressure in mm Hg	Posterior Tibial Artery Pressure in mm Hg	Difference in mm Hg	Height of Column separating Armlets in cm	Difference calculated from Height of Column in mm Hg
H.H.R.	Horizontal . .	140	138	**2**	—	—
	Standing . .	136	204	**68**	89	**68.5**
	L posture, legs up	122	76	**46**	60	**46.1**
	Vertical head down	148	70	**78**	101	**77.7**
P.H.R.	Horizontal . .	126	126	—	—	—
	Standing . .	140	204	**64**	86	**66.1**
	L posture, legs up	132	78	**54**	65	**50.0**
	Vertical head down	116	42	**74**	91	**70.0**

Comparison of the systolic blood pressure in the brachial and posterior tibial artery (M. Flack & L. Hill (1909), Measurement of systolic pressure in man, *Brit. med. J.* i, 273).

Venous Pressure.—The venous pressure in man is best measured directly by connecting a cannula, introduced into a vein, to a manometer filled with a 3 per cent. solution of sodium citrate. Measurement of the actual hydrostatic pressure within a vein, say at the level of the elbow, is of little value since the figure obtained depends on the position of the vein relative to the heart. Thus if the arm is raised the pressure falls and if the arm is lowered it rises. The upper level of fluid in the manometer tube will, however, occupy a constant relation to the level of the heart, or to the sternal angle which is a convenient reference point whose relation to the heart is reasonably constant. Thus if there be no obstruction or constriction between the manometer tube and the right atrium the difference in level between the height of the meniscus and the sternal angle will represent the pressure in the right atrium. Normally this lies between zero and -2 cm of water. Of course, the effective venous pressure for filling the heart is greater than this because of the subatmospheric pressure within the thorax. The latter amounts during inspiration to -5 mm Hg ($\equiv 67$ mm water). Thus the *effective filling pressure* is $6{\cdot}7-(-2)$, or $8{\cdot}7$ cm of water.

The external jugular vein may be made to act as its own manometer, since it has no functioning valves. The vessel collapses at the point at

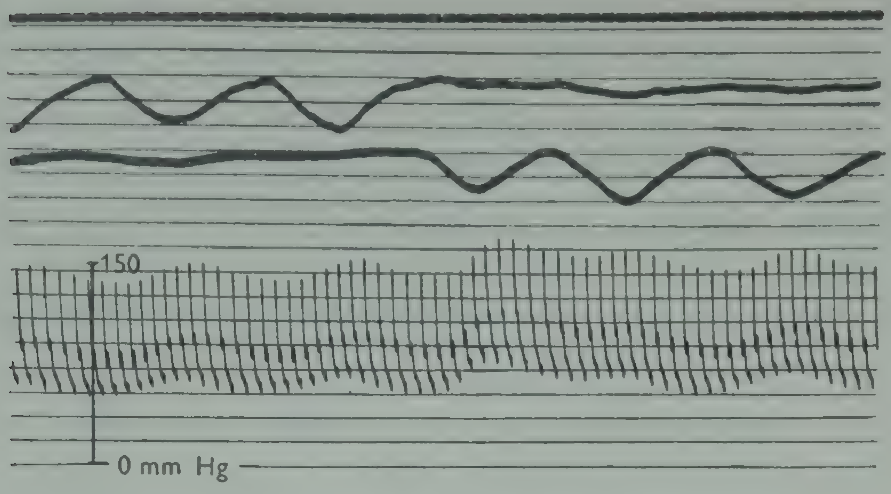


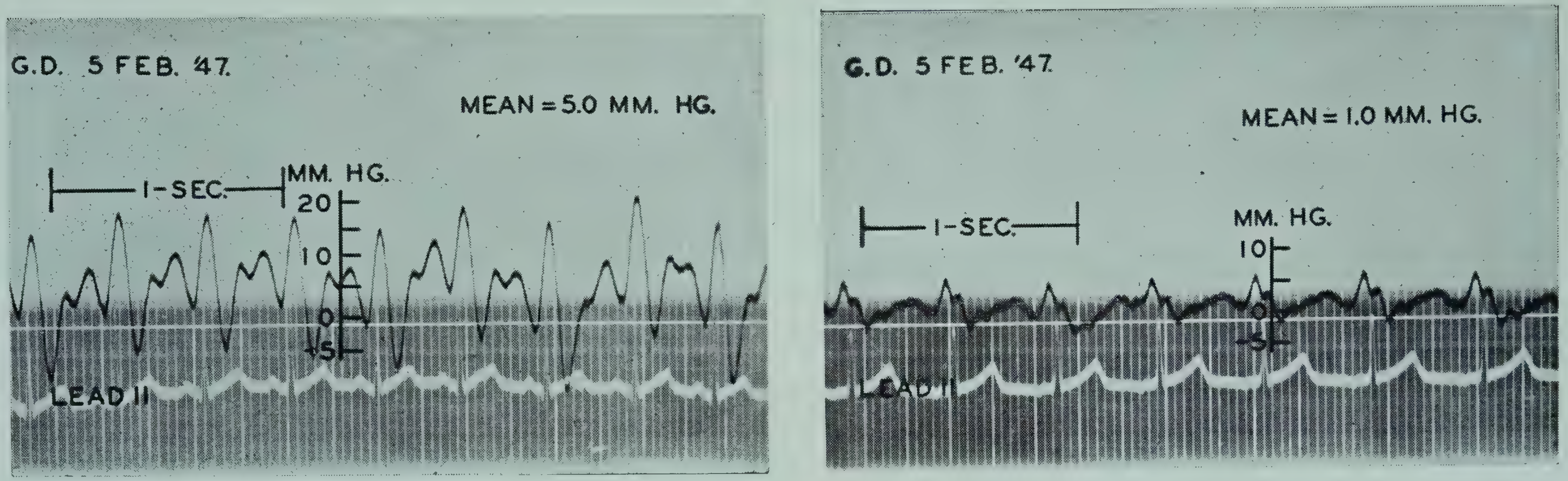
Fig. 29, 11

Tracings in a normal man of, from above downwards, time in seconds, thoracic and abdominal breathing, and arterial blood pressure. The respiratory records were made with spirometers; a downward movement indicates inspiration. The blood pressure was measured with an electrical (capacitance) manometer communicating with the inside of an artery. The arterial pressure shows variations corresponding to respiration which do not alter in phase when the breathing changes from thoracic to abdominal (A. C. Dornhorst, P. Howard & G. L. Leathart (1952), *Circulation*, **6**, 554).

Table 29, 12

Variations in arterial pressure in an apparently healthy female aged 32. This table illustrates the considerable alterations that can occur under ordinary circumstances and emphasizes the care that may be required to obtain a steady reading.

Time Minutes	Arterial Pressure mm Hg		Circumstances
	Systolic	Diastolic	
0	140	90	Patient sitting, immediately after walking into the room for the first time
7	130	80	Sitting talking
10	135	90	Patient lies on couch
25	120	80	After lying quietly on couch for 15 min
30	125	80	After lying quietly on couch for 20 min
35	120	75	After lying quietly on couch for 25 min
40	120	75	After lying quietly on couch for 30 min
41	..	..	Right hand suddenly immersed in ice-cold water for 30 sec
41¼	150	90	The response to the discomfort produced by cold
41½	165	95	
42	140	90	
45	130	85	
55	120	75	The basal blood pressure
55½	140	85	Immediately after an unexpected loud noise is made behind the patient's head
65	140	90	The blood pressure remains elevated
85	125	80	20 min later the pressure is still not basal
85½	125	85	Patient suddenly stands erect


G.D. 5 FEB. 47.
MEAN = 5.0 MM. HG.
1-SEC.
MM. HG.
20
10
0
-5
LEAD II
G.D. 5 FEB. '47.
MEAN = 1.0 MM. HG.
1-SEC.
MM. HG.
10
0
-5
LEAD II

A. Left atrium.

B. Right atrium.

FIG. 29, 13

Two records of blood pressure in a five months' old child under avertin anæsthesia taken within a few minutes. The septal defect which allowed the passage of the catheter into the left atrium was very small and there was no evidence of cardiac decompensation or dilatation of the right or left atria. The rise of pressure on atrial systole, the fall of pressure on diastolic inflow and the mean pressure are all greater on the left side (A. Cournand, H. L. Motley, A. Himmelstein, D. Dresdale & J. Baldwin (1947), *Amer. J. Physiol.* **150,** 268).

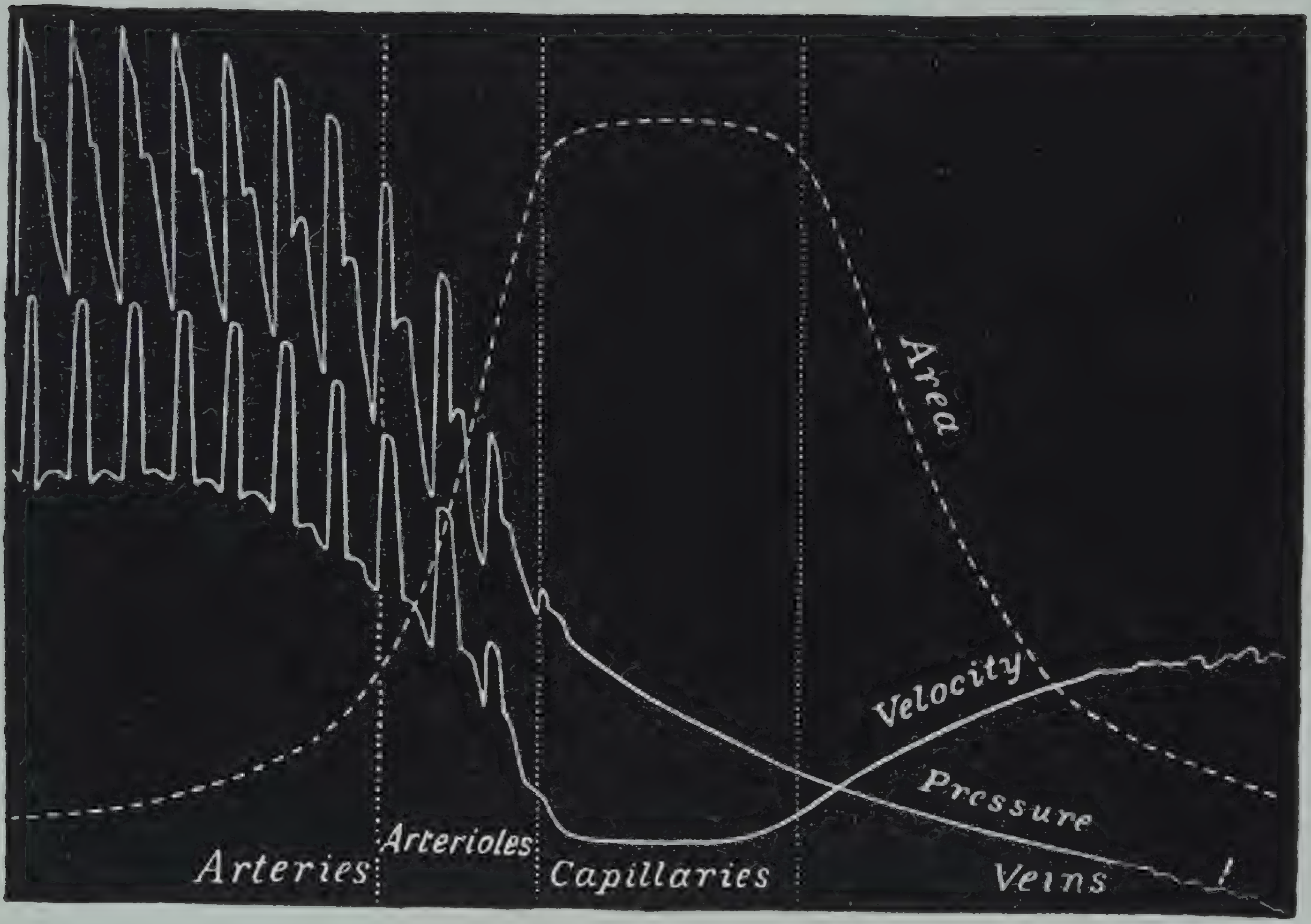


FIG. 29, 14

Diagram showing the relations between pressure, velocity and total cross-sectional area in arteries, capillaries and veins. (*From* L. Fredericq & J. P. Nuel (1920), *Elements de Physiologie humaine*, p. 136. 7th ed. Liège: Vaillant-Carmanne. Partly *after* J. Gad.)

which the external pressure (that is the atmospheric pressure) is greater than the internal pressure (the venous blood pressure). The upper level of the column of blood in the external jugular vein is measured from a horizontal plane passing through the sternal angle (Fig. 29, 15, A) which is taken as the reference point. With the patient reclining and the head supported on a pillow the level is from 0 to 2 cm below this line and the venous pressure is said to be, therefore, 0 to −2 cm (of blood). Higher values are found when the heart is failing or when the blood volume is much increased. A less accurate but sometimes convenient gauge of the venous pressure is obtained by noting the level relative to the sternal angle at which the veins on the back of the outstretched hand collapse.

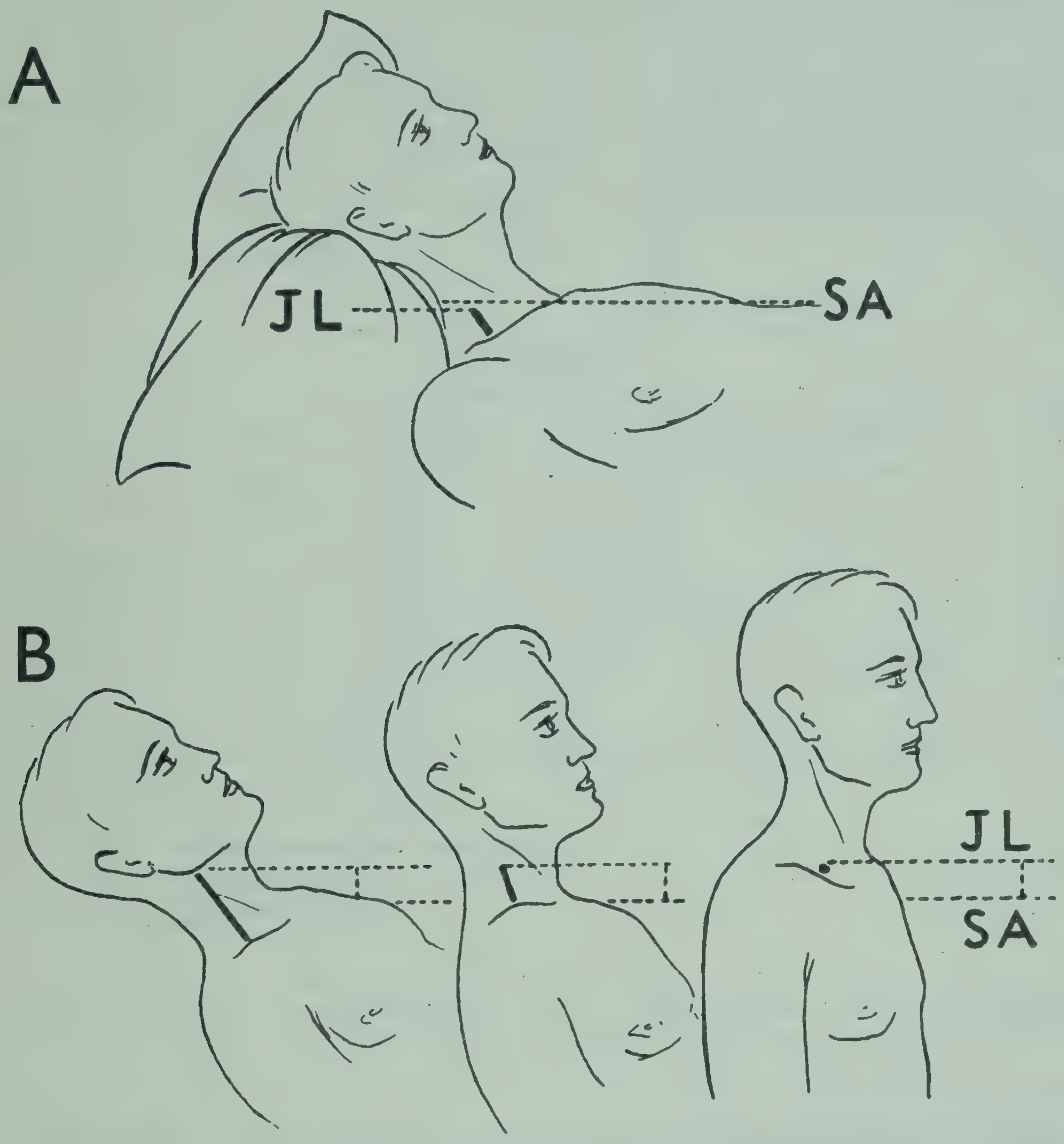


Fig. 29, 15

The upper diagram, A, shows the level of blood in the external jugular vein when the subject is lying with the head supported on pillows. The upper level of the column of blood in the jugular vein (JL) lies a little below the level of the sternal angle (SA). In the lower diagram, B, the upper level of the column of blood in the external jugular vein is above the sternal angle, indicating an increase of venous pressure. Note that as the position of the patient alters the upper level of the column of blood occupies different positions in the neck although its relation to the sternal angle remains the same (T. Lewis (1949), *Diseases of the Heart.* London: Macmillan).

The pressure in the atria and in the great veins is not constant but varies with the cardiac cycle and with respiration (Fig. 29, 13).

The venous pressure in the superficial veins of the dorsum of the foot with the subject erect and motionless is from 70 to 100 mm Hg, which is the same as the calculated hydrostatic pressure from the right atrium. On exercise, owing to the action of the muscle pump (p. 436), the reading falls to less than 30 mm Hg unless the leg veins are varicose with incompetent valves, when there may be little difference between the readings during exercise and at rest.

Capillary Pressure.—Since the circulation of the blood is dependent on the beat of the heart and the elastic recoil of the aorta and large arteries, the pressure in the capillaries must be lower than that in the smallest arteries and greater than that in the smaller veins (Fig. 29, 14). The arteries and the capillaries have considerable powers of relaxation and contraction according to the metabolic needs of the tissues, and thus the pressure in the capillaries varies considerably from time to time. The only satisfactory method of measuring intracapillary pressure was devised by Landis in 1930. He inserted a micropipette of pyrex glass with an aperture of about 10μ into various parts of the capillary loops of the skin of the nail bed. The pipette, moved by a micromanipulator and observed through a binocular microscope, was connected by a system of tubes containing citrate solution to a mercury manometer. When the pipette was inserted into a capillary the pressure was adjusted until blood neither entered nor left the micropipette. This pressure was taken as the internal capillary pressure. The average pressure (Fig. 29, 16) in the arterial limb was 32 mm Hg with a range of 21 to 48 mm Hg; the summit of the loop gave a mean pressure of 20 (range 15 to 32) and the venous end a mean of 12 mm Hg (range 6 to 18). There is thus a considerable gradient of pressure along the capillaries near the root of the nail. The pulse pressure at the arteriolar end of the loop was at least 5 to 10 mm Hg and at the venous end usually zero. When a pneumatic cuff was put round the arm to occlude the venous return the pressure in the arterial end of the loop rose at once to a level about 10 mm Hg above that in the cuff. When the hand was held above the head the venous pressure fell below atmospheric. The veins then collapsed and offered a considerable resistance to flow, thus keeping the capillary pressure gradient normal. When the hand was lowered again the capillary pressure rose almost as much as would be calculated from hydrostatic considerations. Warming the hand produced arterial dilatation and the pressure in the capillary loop rose to

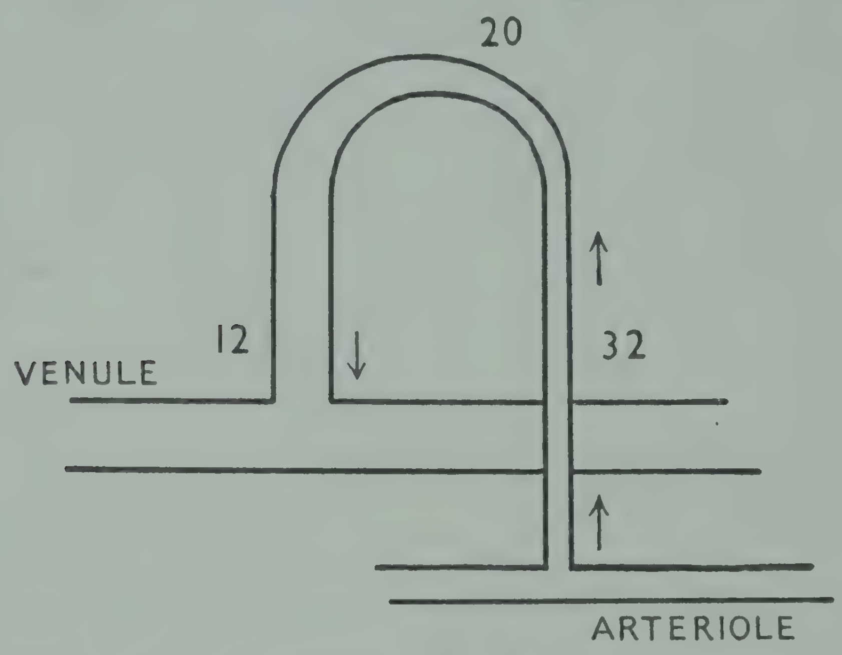


FIG. 29, 16

Average pressures in mm Hg in a skin capillary. (*After* E. M. Landis (1930), *Heart,* **15**, 213.)

20 to 30 mm Hg, whereas cooling the skin resulted in a fall of about 10 mm Hg.

Determinants of Arterial Pressure.—The changes in systolic and diastolic pressures at various points of the vascular system are indicated diagrammatically in Fig. 29, 14, and the mean intravascular pressures in various vessels in Table 29, 17. The biggest fall in arterial pressure occurs in the arterioles which offer a large resistance to flow. The arterial pressure is thus to a considerable extent dependent on the calibre of the arterioles, dilatation leading to a fall of pressure and constriction to an increase. This phenomenon is referred to as the *peripheral resistance* and is considered in more detail on p. 499.

TABLE 29, 17

Table of Mean Blood Pressure at various points in the vascular system in Man (*supine*)

	mm Hg
Large artery (for example carotid)	90
Medium artery (for example radial)	85
Capillary	10 to 30
Small veins of arm	9
Intestinal capillaries	over 10
Portal vein	10
Inferior vena cava	3
Large veins of neck	0 to −2
Pulmonary artery	15

Because of the high resistance to flow offered by the arterioles, the pressure in the capillaries is largely independent of the pressure in the big arterial trunks. The fall of pressure as the blood flows sluggishly through the capillaries and veins is small because these structures offer much less frictional resistance than the arteries. The capillary pressure is determined to a considerable extent by the venous pressure because there is little resistance to flow from capillaries to veins. Since, however, the capillary bed is able to accommodate more than the total blood volume the capillaries are capable of exerting a considerable influence on the level of arterial pressure. The low pressure in the veins is accounted for by the fact that the resistance to flow through the tricuspid valve is normally low. Any increase of resistance to outflow is followed by a generalized increase of venous pressure.

Apart from the resistance of the walls of the vascular system, however, there are, of course, other factors helping to maintain the arterial pressure. The output of the heart for example is a second important factor. This in itself depends upon an adequate venous return to the heart and upon the heart rate, for at very high rates of contraction the output per minute will be quite small because of the shortness of each successive diastole. Thirdly, the cardiac output depends upon the heart muscle; when this is grossly diseased, the output is reduced and the arterial pressure falls.

The level of arterial pressure is also governed by the amount of blood in the system and by its viscosity. Rapid loss of blood from the circulatory system, especially if it be from an artery, leads to a fall in arterial pressure; overloading of the circulation to a rise. An increase in the viscosity of blood which is sometimes met with in disease is often accompanied by an elevated arterial pressure.

Finally the difference between the systolic and diastolic pressure (pulse pressure) is determined by the characteristics of the walls of the vessels. When the walls of the large arteries become rigid and inelastic they are the less easily able to accommodate the systolic discharge of the left ventricle and there is an undue increase in pressure at each systole (systolic hypertension).

The main factors governing arterial blood pressure are summarized in Fig. 29, 18.

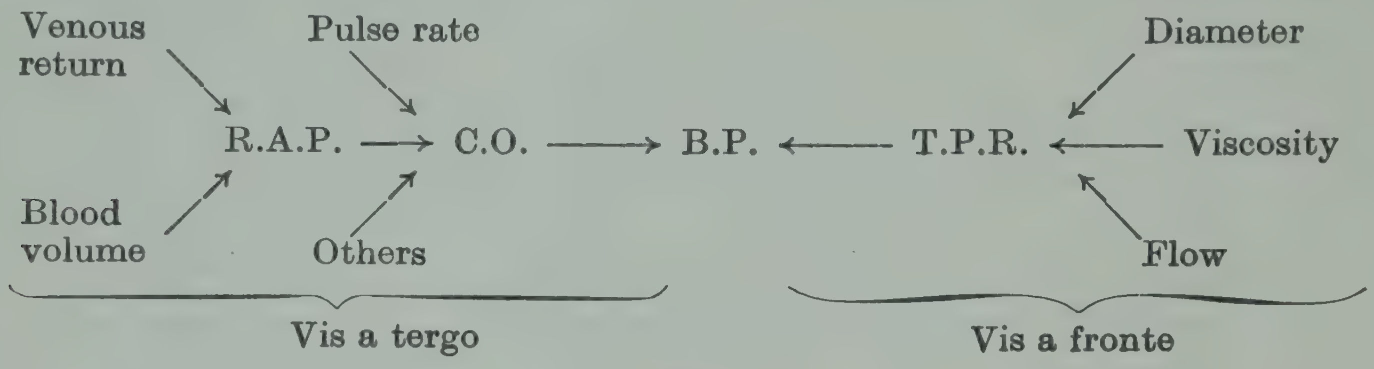


FIG. 29, 18

Diagrammatic summary of the factors governing blood pressure.

R.A.P. = right atrial pressure.
C.O. = cardiac output.
B.P. = blood pressure.
T.P.R. = total peripheral resistance.

The Effect of Gravity upon the Circulation.—The effect of gravity on the circulation has already been mentioned on pp. 486 and 487, where the pressure at different parts of the arterial tree was compared in the erect and supine position. The effect of gravity upon the venous pressure has also been discussed on page 489. These effects are, however, quite small compared with those encountered in aircraft which can perform tight turns. In these circumstances the acceleration may be several times g ($g = 981$ cm/sec/sec). If in such a turn the acceleration is $5g$ all the tissues of the body weigh five times as much as normal and, in the case where the head is nearer the centre of rotation than the pelvis, all those parts which are not firmly fixed, for example the tissues of the face, sag. The blood weight also increases and the blood tends to leave the head so that the carotid pressure falls. When the carotid pressure falls below 50 mm Hg there is a loss of contrast over the visual field and, if the carotid pressure falls below the intraocular pressure blood cannot enter the eye and vision is totally (but temporarily) lost. A slightly greater fall of blood pressure produces unconsciousness. The actual value of g at which consciousness is lost is dependent on the rate of acceleration. If the peak acceleration is reached slowly there may be time for the carotid sinus response to occur. This happens in about 10 seconds and would, of course, have the effect of preventing a fall in carotid blood pressure. The acceleration required to produce loss of vision is quite variable; for example, if the time of exposure to the peak acceleration is 5 seconds, loss of vision may occur at anything from $3g$ to $9g$. The reason for this and other variations among fit young men in their responses to acceleration is not known.

A period of unconsciousness produced by a tight turn in an aircraft, or experimentally in a centrifuge, may be followed by a period of convulsions during the restoration of blood flow in the cerebral vessels, that is at a time when reoxygenation is occurring. The cause of the convulsions is not clear but they may be related to changes in cerebrospinal fluid pressure. During acceleration the cerebrospinal fluid pressure in the cranium is much reduced but the pressure within the spinal theca is much increased.

All the methods of attempting to counteract the effect of *g* are, of course, aimed at maintaining the arterial pressure in the head. For acrobatic flying during the war a crouching position was adopted by raising the rudder bar pedals; this reduced the difference in level between the head and heart and lessened the pooling of blood in the legs. This position quite markedly increased the resistance to blackout. Anti-*g* suits applied to compress the legs and thighs as well as the abdomen are also beneficial by preventing pooling of blood below the level of the heart.

THE ARTERIAL PULSE

The Cause of the Pulse.—The pulse wave in the arteries is dependent on three factors: (1) the *intermittent inflow* of blood from the heart, (2) the *resistance* to outflow of blood from the arterioles into the capillaries and (3) the *elasticity* of the arterial walls. The phenomenon may be readily understood by analogy with the action of a stirrup-pump. At each stroke of the pump (systole) a pulse passes down the hose-pipe and at the same time the elastic hose-pipe is distended because the fine nozzle at the end prevents free exit of water, that is to say, it offers a resistance to outflow. Between the strokes of the pump (that is in diastole) the hose-pipe contracts because it is elastic and continues to force water through the nozzle, so that if the rate of pumping is not too slow a steady stream of water flows from the nozzle. And so it is in the circulation. The pulse wave is almost non-existent in the capillaries unless the arterioles are dilated. In the veins the rate of flow is uniform at rest except in the large veins near the heart.

The rapid rise of pressure following the ejection of some 60 ml. of blood from the left ventricle at each systole expands the aorta and a pressure wave or pulse passes along its wall. The pulse is *not* due to the passage of blood along the arteries—the blood travels at only 0·5 metres per second—but to a pressure wave travelling about 7 metres per second. If the piston P in Fig. 29, 19, is driven towards the right in the tube T tied into an artery the vessel is distended locally, because the blood is incompressible. This produces a local increase of pressure that has little effect on the advancing piston, which is in fact a mass of incompressible blood; instead the next section of the artery is stretched so that a wave of pressure travels along the vessel wall without involving actual transmission of blood. Indeed the pulse wave can travel down such an elastic tube even if the distant end is closed; there it may be reflected back in an attenuated form. Reflexion of pressure waves occurs at every branching point in the arterial tree and makes records of the arterial pulse very complex.

Pulse-wave Velocity.—The speed of the wave depends on the elasticity of the vessel walls and on the inertia of the blood. Although the inertia of the blood remains the same the velocity of the pulse wave increases with age because the arteries tend to harden. Thus if the pulse is recorded simultaneously from the carotid and the radial arteries the velocity of the pulse wave is found to be about 5 metres per sec at 5 years and 8 metres per sec at 60 years. The pulse wave does not occur simultaneously at all parts of the arterial tree; in the radial artery it occurs later than in the carotid and the pulse in an artery in the foot is later still.

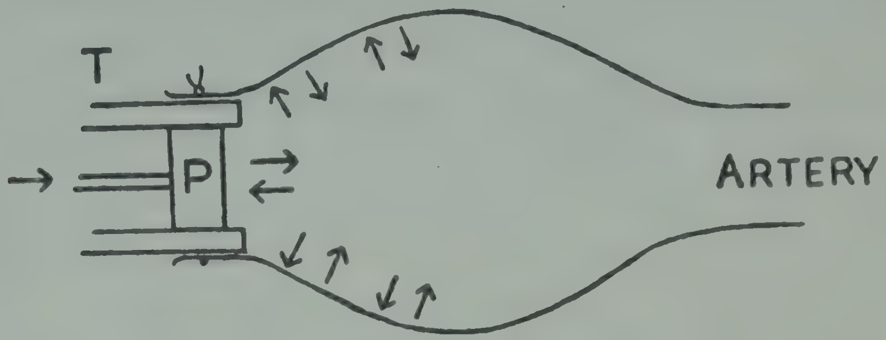


Fig. 29, 19

Diagram to illustrate the initiation of a pulse wave in the aorta.

It is useful to know that the pulse wave arrives at almost the same moment at the beginning of the femoral artery and at the radial artery at the wrist. If the femoral pulse is later than the radial as judged by palpation it is probable that the wave is reaching the femoral artery along collateral channels developed as the result of a congenital narrowing of the aorta (coarctation).

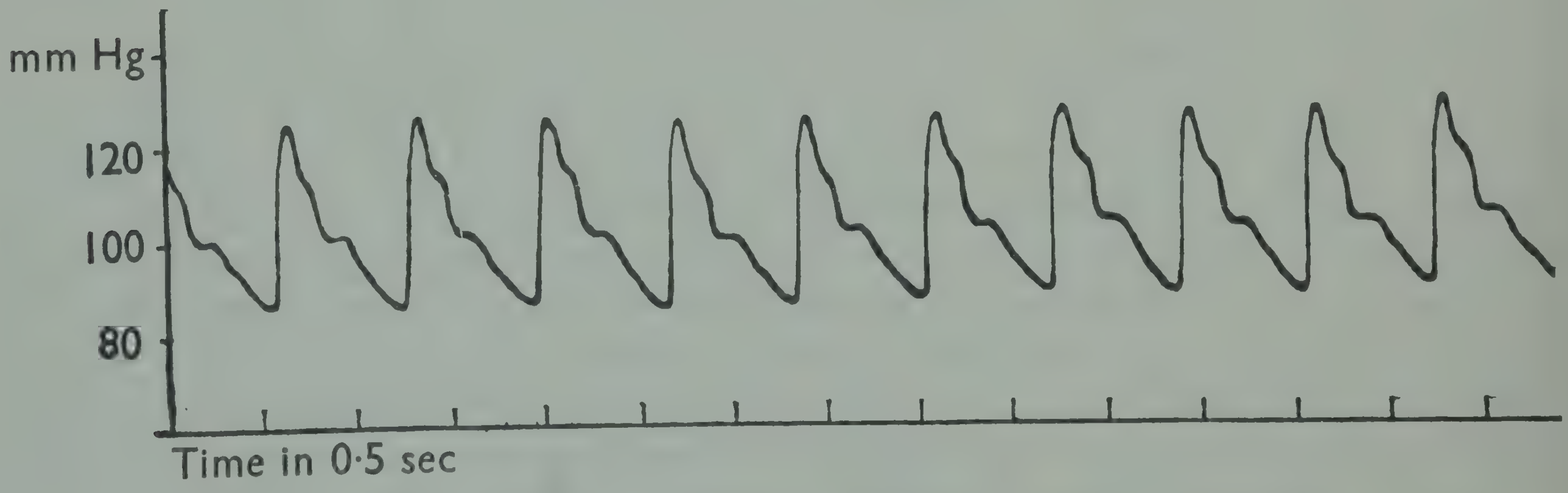


Fig. 29, 20

Continuous trace of pulse waves in the brachial artery of a healthy young man recorded with a transducer manometer in communication with the lumen of the artery.

(*By courtesy of G. Cumming.*)

The Shape of the Pulse Wave.—The shape of the pulse wave is nowadays studied in two principal ways. The first uses a pressure sensitive capsule applied to the artery and records only the shape of the pulse wave, without in any way measuring the pressures involved. The second, which makes use of an electronic manometer communicating through a needle with the inside of the artery, is capable of measuring the actual pressures as well as recording the shape of the pulse wave. Modern tracings are, generally speaking, free from the instrumental artifacts produced by earlier recorders and are probably accurate records of the form of the pulse wave. Fig. 29, 20 is a recording of the pulse wave in the brachial artery. This consists of a steeply ascending limb, due to ventricular systole, followed by a more gradual fall in pressure as the blood drains away through the peripheral vessels. This diastolic limb should thus be dependent on the peripheral resistance. However, the curve is not a smooth one, as would be expected if it were merely a slow leak away of pressure, but contains several minor oscillations. The most constant of these is the dicrotic wave, which is the largest wave on the descending line of the trace. It was formerly supposed that the dicrotic wave is caused by a rebounding of blood from the closed aortic valves but it is more probable that this wave, together with other smaller subsidiary waves often observed, represents oscillations of the aorta at its own natural frequency. The pulse wave in the aorta has already been described on p. 433, but the form of the pulse wave is changed as it passes to the periphery by the frictional resistance of the arterial walls and by reflected waves, so that even with an ideal recording device interpretation of the wave form is difficult.

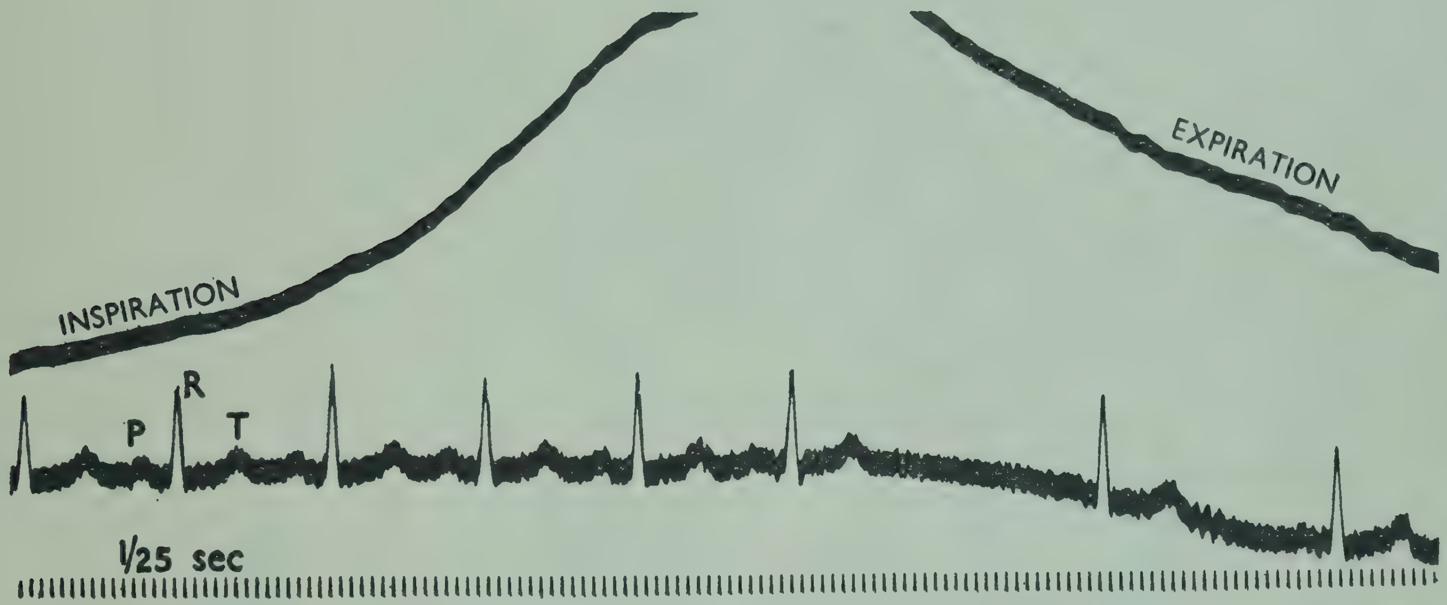


FIG. 29, 21

Sinus arrhythmia. Electrocardiogram (lead II) from a healthy male aged 14 showing cardiac acceleration during inspiration and slowing during expiration. Time trace, $\frac{1}{25}$ sec.

With the surge of blood into a limb at each heart beat there is a momentary increase in volume of the muscles. This can readily be observed in a person sitting with his legs crossed; the raised foot gives a little kick forward with each pulse beat as the calf muscles expand. This volume pulse occurs a little later than the corresponding arterial pulse. It can be recorded by putting the limb in a rigid box, or plethysmograph, connected with a tambour the movements of which are recorded by a mirror or a lever (Fig. 29, 25).

The Pulse Rate.—The pulse rate varies considerably from individual to individual and in the same subject at different times, depending on a large number of factors which are considered on p. 467. The resting pulse rate tends to fall as age advances, as shown in Fig. 28, 43.

Sinus Arrhythmia.—In adult life, so far as can be judged by the fingers, the pulse is regular, but accurate records show that there is a slight acceleration during inspiration and a slight deceleration during expiration. This is

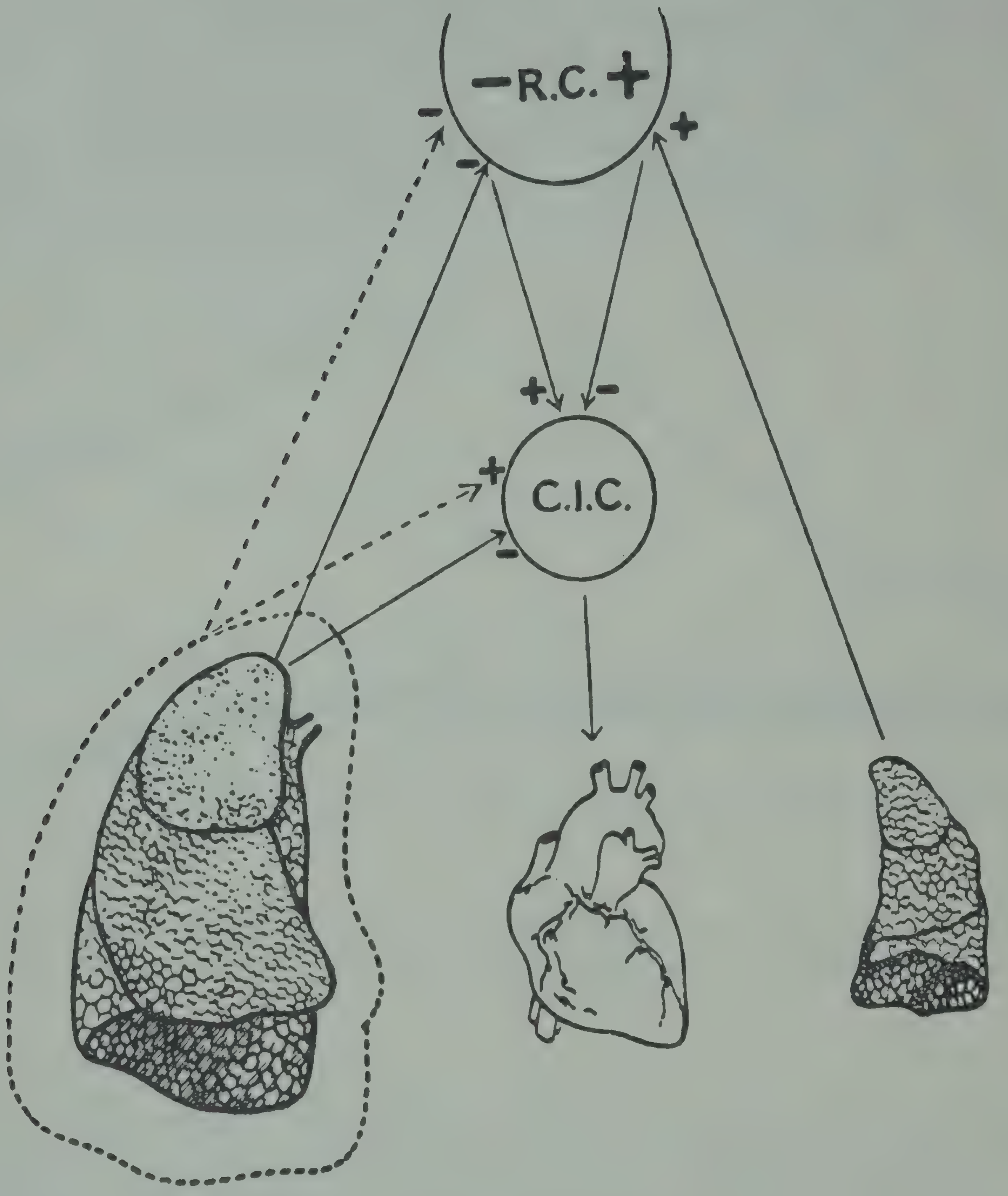


FIG. 29, 22

A diagrammatic representation of the interrelations between the respiratory act and the heart rate. R.C., the respiratory centre; C.I.C., the cardio-inhibitory centre; + denotes excitation; − denotes inhibition. On the right-hand side the collapse of the lungs is shown to induce an inhibition of the cardio-inhibitory centre brought about indirectly through exciting the respiratory centre. On the left-hand side is shown an inflated lung which causes a direct inhibition of the cardio-inhibitory centre and simultaneously an indirect excitation of it brought about by inhibiting the respiratory centre. The dotted contour of the lungs illustrates an over-expanded lung which causes the opposite effect, namely, a direct excitation of the cardio-inhibitory centre accompanied by an inhibition of the respiratory centre. These latter effects have nothing to do with the respiratory arrhythmia (G. B. Anrep, W. Pascual & R. Rössler (1936), *Proc. Roy. Soc.* B, **119,** 228, Fig. 7).

termed *sinus arrhythmia.* In children this fluctuation in rhythm is much more marked (Fig. 29, 21) and is often obvious on palpation of the radial artery. During inspiration the thoracic pressure falls while the abdominal pressure rises and, therefore, blood flows more readily into the thoracic veins, producing a rise of atrial and central venous pressure. This rise produces reflexly an acceleration of the heart (Bainbridge's reflex). This is only a partial explanation of sinus arrhythmia, since alterations of heart rate related to respiratory activity can occur independently of any alteration in venous pressure. The central and reflex causes of respiratory arrhythmia have been analysed by Anrep and a summary of his findings is given below.

The causes of respiratory arrhythmia are partly central and partly reflex. The reflex mechanism can be studied separately if the respiratory centre is put out of action by supplying it with blood of low carbon dioxide content. In these circumstances inflation of the lungs produces cardiac acceleration; since the acceleration does not occur if the vagus nerves are cut immediately above the roots of the lungs (care being taken to avoid damage to the cardiac fibres), inflation must stimulate sensory nerve endings in the lungs. After vagus section at this level the heart is slowed. This shows that the lungs send a constant stream of impulses up the vagus nerves; during inflation the frequency of the impulses increases and the heart accelerates. The endings show adaptation because a prolonged inflation results in an immediate acceleration, followed by a slowing of the heart. In these experiments, in which the respiratory centre is in abeyance, further investigations show that the sympathetic cardio-accelerator mechanism is not involved. The impulses arising in the lungs must, therefore, travel up the vagus nerves direct to the cardio-inhibitory centre to inhibit it.

The central mechanism of respiratory cardiac arrhythmia can be studied by abolishing the pulmonary reflexes by section of the vagi in the thorax just above the lungs. When the carbon dioxide tension in the blood is adequate the respiratory centre discharges rhythmically, sending impulses down the phrenic nerve to bring about inspiration by contraction of the diaphragm; cardiac acceleration occurs when the respiratory centre is sending out impulses. If the carbon dioxide content of the blood is reduced the respiration is weakened and respiratory cardiac arrhythmia is diminished. At the same time as impulses pass out of the respiratory centre to the phrenic nerves other impulses pass to the vagal cardio-inhibitory centre to inhibit it, and so produce cardiac acceleration. The sympathetic cardio-accelerator centre acts in a reciprocal sense to the cardio-inhibitory centre, since it is possible to demonstrate respiratory arrhythmia when the cervical vagi are cut if the heart is made to beat slowly by stimulation of the peripheral end of one vagus. The cardio-inhibitory centre by alterations of vagal tone has, however, the greater influence on the heart rate. Sympathectomy scarcely affects respiratory arryhthmia.

Correlation of Reflex and Central Respiratory Arrhythmia.—Inflation of the lungs produces an inhibition of the respiratory centre (Hering-Breuer reflex, p. 562) and the phrenic discharge to the diaphragm ceases and expiration begins. This causes a reduction of the impulses mentioned in the previous paragraph, which pass, during the discharge of the respiratory centre, to the cardio-inhibitory centre. Inflation of the lungs thus, indirectly through the respiratory centre, produces a slowing of the heart. But, as already mentioned (p. 496), inflation of the lungs causes a reflex acceleration of the heart directly through the cardio-inhibitory centre.

The endings responsible for the cardio-accelerator effect are less robust than those responsible for the Hering-Breuer reflex. In pulmonary œdema and after overstretching of the lungs by excessive inflation the Hering-Breuer reflex alone remains, and inflation of the lungs produces a slowing of the heart.

The diagram (Fig. 29, 22) sums up Anrep's analysis of respiratory arrhythmia. The vagal cardio-inhibitory centre is inhibited (1) by the respiratory centre when it is discharging, and (2) by impulses arriving from the lungs when these are inflated. The vagus centre during inflation is affected indirectly in the opposite direction by the impulses which inhibit the respiratory centre (Hering-Breuer reflex). Deflation of the lungs has the opposite effect. The central regulation of respiratory arrhythmia is easier to demonstrate than the pulmonary reflex and is probably the more important. To complete the picture for the intact animal the effect of the right atrial reflex (p. 466) must be added, although it is probably of minor importance.

VENOUS PULSE

The flow of blood in the veins at rest is continuous except near the heart. The variations of atrial pressure usually pass back a short way into the veins since the large veins near the heart are not equipped with valves. A venous pulsation at the root of the neck is therefore often seen in normal people. The venous pulse in man can be recorded by placing a metal cup on the right side of the neck above the sternal end of the clavicle. A rubber tube leads from the cup to a tambour attached to a lever the movements of which can be recorded on a moving paper or a photographic plate. Such a record is shown in Fig. 29, 23; it cannot be analysed properly unless either an electrocardiogram or a record of the arterial pulse is made simultaneously.

In the venous pulse there are three positive waves during each cardiac cycle. The *a* wave is produced by atrial systole. It is replaced by a rapid fluttering movement when the atria flutter (p. 477) and disappears when they fibrillate (p. 477). The *a* wave is followed shortly by the *c* wave. This marks the closure of the tricuspid, and therefore of the mitral, valve, and is synchronous with the isometric phase of ventricular systole. It is accompanied by, and prolonged into, a disturbance imparted to the recording apparatus by the pulse in the underlying carotid artery. Some time later the *v* wave appears. This is due to the gradual increase of venous pressure during ventricular systole and the summit of the *v* wave is coincident with the opening of the mitral and tricuspid valves. The first negative wave, between *a* and *c*, is the result of atrial diastole. The trough between *c* and *v* is produced by descent of the heart during ventricular systole and the third negative wave (after *v*) is due to the sudden release of pressure in the veins as the tricuspid valve opens.

The tracing of the jugular pulse in Fig. 29, 23, should be compared with the record of the changes in intra-atrial pressure obtained from a membrane manometer communicating directly with the interior of the right atrium (Fig. 28, 28 and Fig. 29, 13).

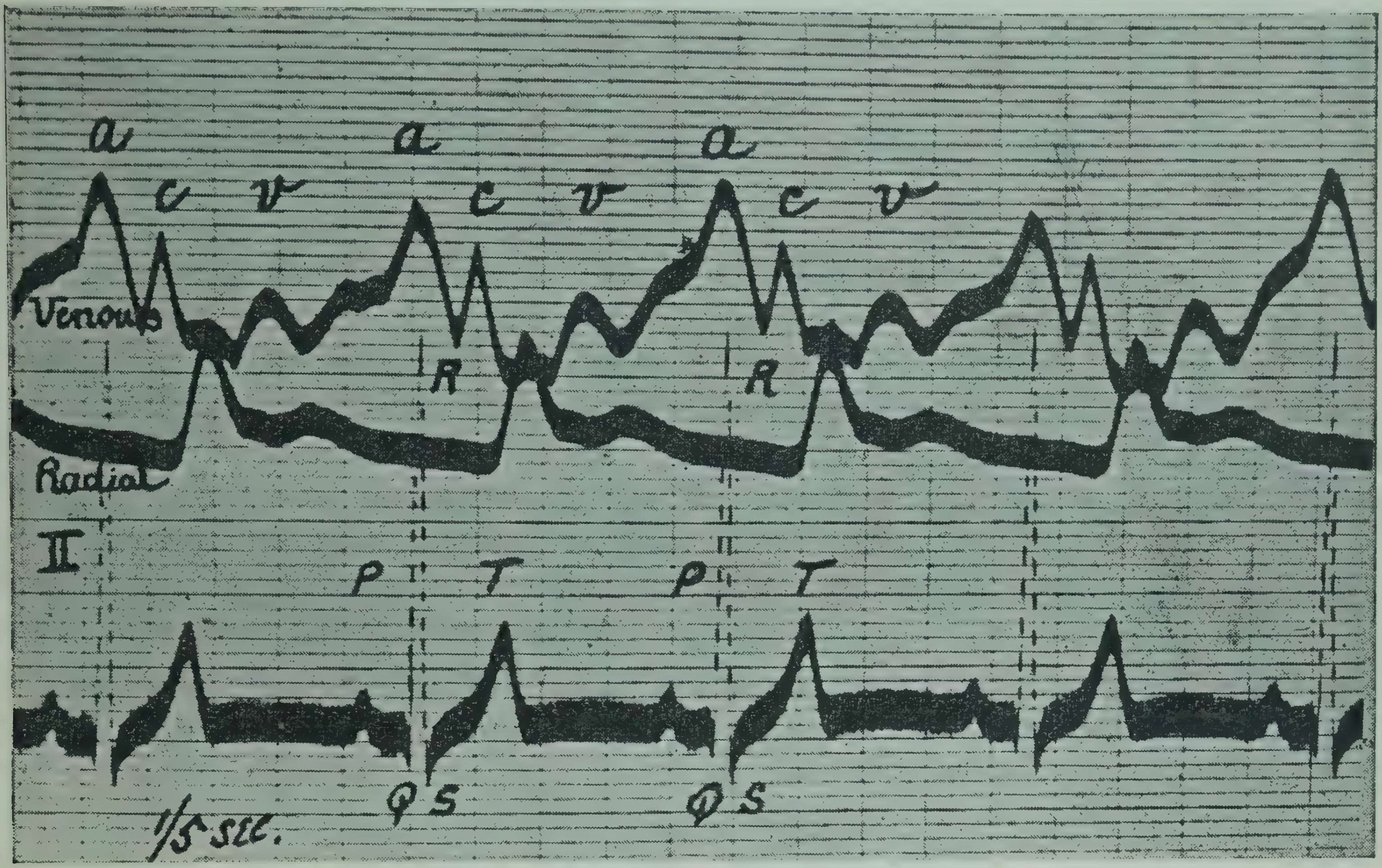


Fig. 29, 23

The upper trace is the jugular pulse obtained by the method described opposite. The lever attached to the tambour cast a shadow on the slit of the recording camera. The radial pulse (middle trace) was obtained by connecting a tambour on the pulse to a tambour with a lever in front of the camera slit. The lowest trace is a string galvanometer record of the electrocardiogram (T. Lewis (1925), *The Mechanism and Graphic Registration of the Heart Beat.* 3rd ed. London: Shaw & Sons).

PERIPHERAL RESISTANCE

The mean arterial pressure depends on the amount of blood which enters the arterial system from the heart (cardiac output) and the rate at which blood leaves the arterioles to pass into the capillaries and veins, that is, on the resistance to the outflow of blood from the arteries, or *peripheral resistance*, perhaps better called the pressure-flow ratio. The relation between these three factors can be expressed thus:

$$\text{Peripheral resistance} \propto \frac{\text{Mean arterial pressure}}{\text{Cardiac output}}.$$

Both mean arterial pressure and cardiac output can now be measured in man with considerable accuracy. The pressure can be expressed in dynes per sq. cm by multiplying the reading in mm Hg by 1332; the output is expressed in ml. per sec. Then,

$$\text{Peripheral resistance} = \frac{\text{Mean arterial pressure in mm Hg} \times 1332 \text{ dynes cm}^{-2}}{\text{Cardiac output cm}^3\text{ sec}^{-1}}$$

$$= \text{Dynes sec cm}^{-5}$$

For example, if we take the mean arterial pressure at rest as equal to the diastolic blood pressure plus one third of the pulse pressure (a reasonable approximation to the figure actually found by an intra-arterial manometer), say $80 + \frac{1}{3}(110 - 80)$ or 90 mm Hg, and the average resting cardiac output as 5 l. per minute or 83 ml. per second, then,

$$\text{Peripheral resistance} = \frac{90 \times 1332}{83} = 1445 \text{ dynes sec cm}^{-5}$$

The total peripheral resistance in man is normally about 1300 dynes sec cm^{-5}.

The layer of fluid immediately in contact with the wall of the artery can be regarded as being stationary. The layer of fluid just inside this stationary layer slides over the latter, and so on in a series of concentric layers, each moving faster than the more peripheral one (laminar flow). The resistance to flow due to this moving of one layer of fluid past another is present even when the arteries are wide and straight. It varies with the viscosity of the blood and the size of the vessel. In 1842 the French physician, Poiseuille, studied the factors governing the laminar flow of viscous fluids through capillary tubes and Poiseuille's Law is an expression of the relationship between these factors.

$$\text{Rate of flow} = \frac{P_1 - P_2}{8l} \times \frac{\pi r^4}{\eta}$$

where P_1 and P_2 are the pressures at the ends of the vessel of length l and radius r and where η is the viscosity of the blood. In other words, if a blood vessel constricts so that its initial diameter is reduced by one half, the same pressure difference across its ends produces only one-sixteenth of the flow. A somewhat greater effect than this is obtained in practice, since the apparent viscosity of the blood is increased by vasoconstriction; the change of apparent viscosity may account for one-third of the total change of resistance to blood flow caused by vasoconstriction.

As the rate of flow of a fluid in a tube is gradually increased there comes a point at which laminar flow is replaced by turbulence and the conditions mentioned above no longer apply. The development of turbulence was shown by Reynolds to be determined by the expression $\frac{rVd}{\eta}$ known as the Reynolds' number, in which r is the radius of the tube, V the velocity of flow, d the density and η the viscosity. When this number exceeds a certain value the flow becomes turbulent. It is obvious then that turbulent flow occurs more readily when the fluid is of low viscosity and the velocity relatively great and that it tends to develop first in large vessels. These conditions are found in the heart and aorta of anæmic patients and the turbulence no doubt accounts for the hæmic murmurs readily audible with the stethoscope over the heart and large arteries in such people. The relatively high velocity of the flow of blood through narrowed valves in the heart, for example, may be sufficient in itself to cause turbulence and the development of audible murmurs (organic murmurs). Turbulent motion in the blood greatly increases resistance to flow.

It can be imagined that in a liquid there are tubes of flow the boundary surfaces of which are formed by stream lines. In such a 'stream tube' the rate of flow or mass crossing every normal section per unit time is constant and the energy in the tube remains constant. From Bernouille's theorem this energy can be expressed as the sum of the kinetic energy, the potential energy and the energy due to pressure or:—

$$\tfrac{1}{2}mv^2 + g\rho h + p = \text{constant}$$

Thus if the motion occurs under the action of gravity alone and external forces are neglected, the velocity increases as the pressure is lowered and *vice versa*. At a constriction in the tube the pressure is lower than in the straight portion because the velocity is greater and conversely if the tube has a bulge the pressure in the bulge is greater than in the straight portion. In a dilatation of the arterial tree such as an aortic aneurysm the increased area of cross-section at the dilatation reduces the velocity of flow and so increases the pressure at the site of the aneurysm.

SPEED OF THE CIRCULATION

It is possible to think of the velocity changes in the circulation in three different ways: (1) as the volume of blood which flows through the vessels of an organ or tissue in a given time (*blood flow* or *volume flow*), (2) as the linear velocity with which blood passes some fixed point in the circulatory system (*blood velocity*), (3) as the time taken for blood to pass from one point to another of the circulatory system (*circulation time*).

1. **Blood Flow.**—The simplest method of determining the blood flow through an organ is to collect and measure over a given period of time all the blood emerging from the veins. Alternatively the volume of blood entering by the artery may be measured by a *stromuhr*. The most familiar form of this instrument is that devised by Ludwig and illustrated diagrammatically in Fig. 29, 24. This has two bulbs of equal capacity—A and B—mounted on a rotatable platform; bulb A is filled with oil and bulb B with blood. The artery is divided and the two ends are tied over E and F. When the clip on the artery is released the oil is pushed over into B. When B is full of oil the bulbs are rotated by hand on the platform C so that the starting position is regained. If the volume of A is known, then the number of times the stromuhr is rotated in a given time gives the blood flow. Many ingenious stromuhrs have been devised for animal work but they are seldom applicable to man. In Rein's stromuhr, heat is applied to an artery in the living animal by a coil situated midway between two thermocouples placed upstream and downstream on the artery. If the flow is fast the temperature difference between the two couples is small, if slow the difference is large. The apparatus is calibrated by noting the galvanometer deflexions at various known rates of flow through an excised artery of the same size.

The changes of volume of a limb or part of a limb can be recorded by means of a *plethysmograph* (Fig. 29, 25). This consists of a rigid box which can be slipped over the part so as to make a closed system in which small changes of volume can be recorded. If the veins are obstructed for a *short* time the record shows an increase in volume which indicates the flow of blood into the part during the period of obstruction (Fig 29, 26).

In animal experiments this plethysmographic method can be applied to the measurement of the blood flow through an organ such as the kidney. If the venous outflow is temporarily obstructed the kidney swells and displaces air from the rigid box enclosing it into a volume recorder which indicates the flow through the organ.

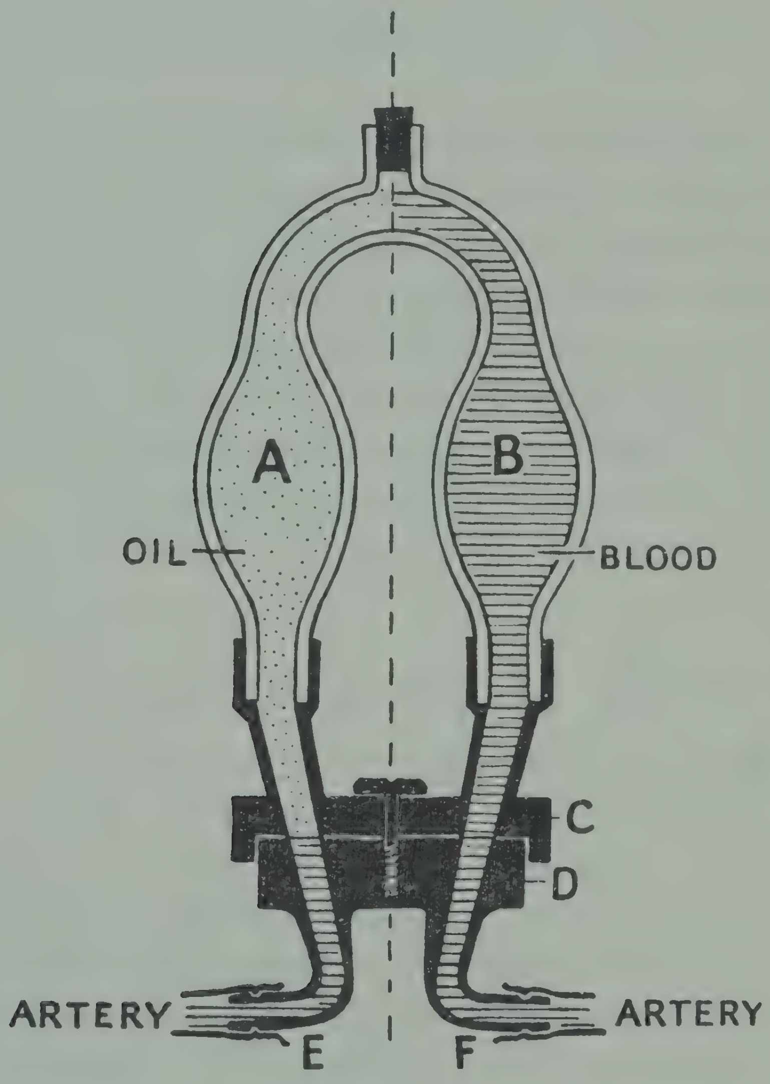


FIG. 29, 24

Ludwig's stromuhr for measuring blood flow.

Since the heat lost from a part is directly related to its blood flow it is possible to assess the latter in a hand or a finger in which the heat loss is high by measuring the heat gained by a water-calorimeter in which the part is immersed. Since the temperature of the skin is determined by the rate of blood flow through it, measurements of surface temperature may be used as an index of blood flow in superficial tissues. Finally the Fick principle (p. 449) can be applied to the determination of the blood flow through an organ, for example the kidney (p. 607) or the brain.

The blood flow through various organs is given in Table 29, 27. These values are approximate only, since they depend upon the state of activity of the part.

2. **Blood Velocity.**—The speed (cm/sec) at which the blood flows through an artery obviously varies from moment to moment if the flow is pulsatile. It is, therefore, convenient to speak of the mean blood

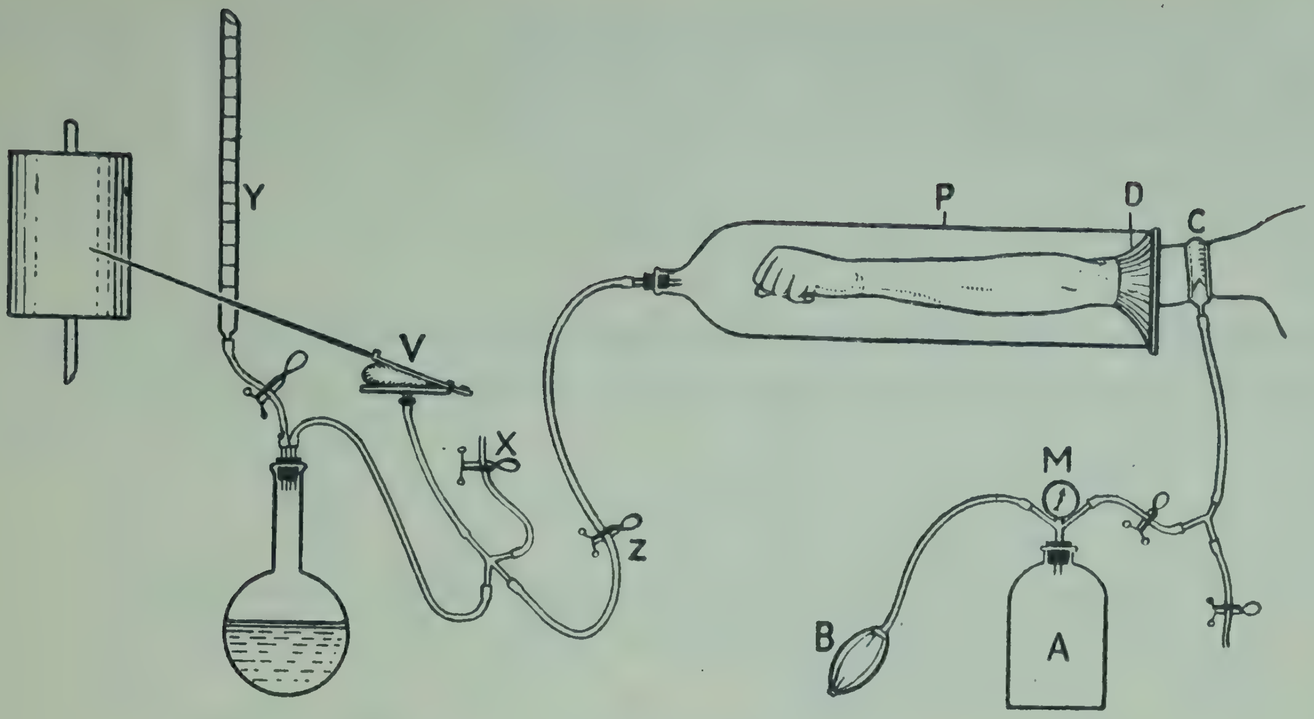


FIG. 29, 25

Diagram of apparatus for determining the rate of blood flow through the arm. The arm is placed in the plethysmograph, P, the opening of which is closed by a piece of rubber dam, D, and the connexion with the skin made tight with soapsuds. The narrow pressure cuff, C, is placed round the arm about 3 cm above the opening into the plethysmograph. The pressure cuff is inflated by opening the stopcock connecting it with the large bottle, A, in which the pressure has previously been raised by the rubber bulb, B. Pressures are read by the spring manometer, M. The plethysmograph is connected with the volume recorder, V, which writes upon a moving drum. Air can be let out of the system by the stopcock, X, and water can be introduced from the burette, Y, so that the writing point of the volume recorder can be adjusted at will. The stopcock, Z, serves to disconnect the plethysmograph from the recording apparatus during adjustments of the former. The recording apparatus is graduated by allowing 5 ml. of fluid at a time to flow in from the burette, and marking the elevation of the volume recorder thus produced (A. W. Hewlett & J. G. van Zwaluwenburg (1909–10). Rate of blood flow in arm, *Heart*, **10**, 88, Fig. 1).

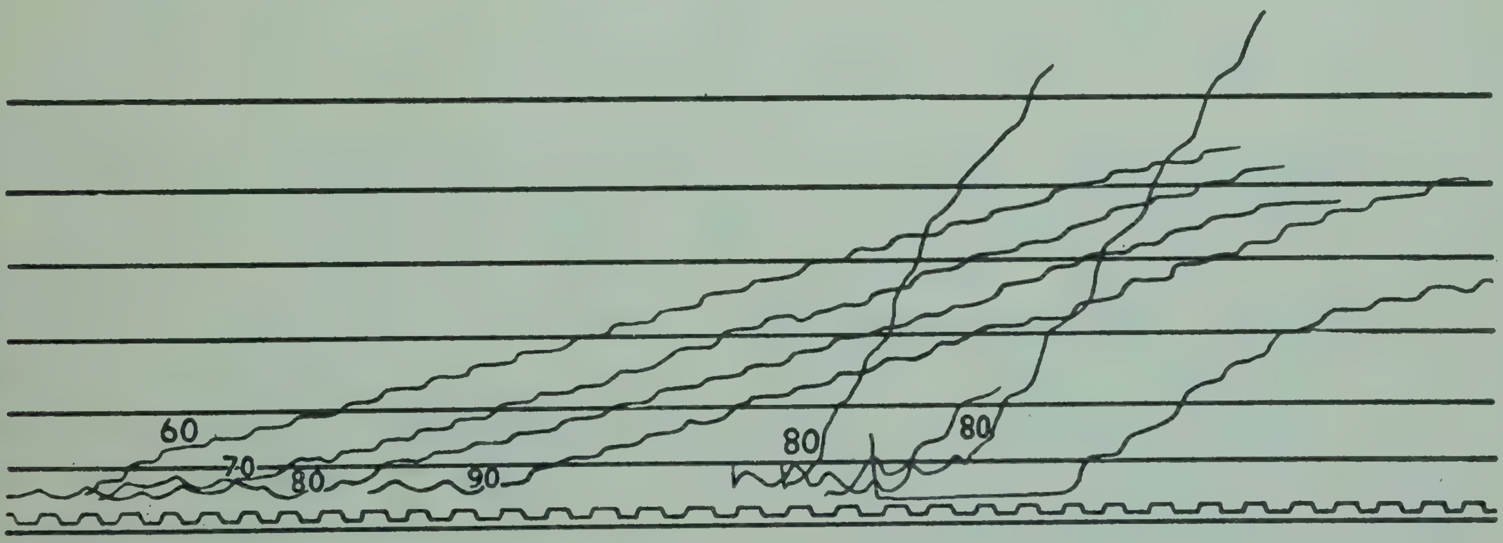


FIG. 29, 26

Plethysmograph records. The first records were made during rest; the following ones at short intervals after exercising the hand within the plethysmograph. The figures on the graph give the pressure in mm Hg in cuff C of Fig. 29, 25. Each space between horizontal lines represents 5 ml. increase in arm volume. Time mark in seconds (A. W. Hewlett & J. G. van Zwaluwenburg (1909–10), *Heart*, **10**, 92).

velocity (V) which can be calculated from the volume of blood (Q) passing in a given time and the cross-sectional area (A) of the vessel in which it is flowing. The relationship is expressed by the equation:

$$V = \frac{Q}{A}.$$

The mean velocity of flow in the aorta calculated from the output of the heart and the cross-sectional area of the aorta is about 40 cm/sec. Since the cross-sectional area of the veins is about twice that of the aorta, it follows that the rate of flow along the veins is about 20 cm/sec. The cross-sectional area of the capillary bed is perhaps a thousand times that of the aorta, so that the rate of flow in the capillaries may be only 0·4 mm per second, which is approximately the value obtained by observation with a microscope of the rate of progression of a red cell along a capillary.

TABLE 29, 27

Approximate values for blood flow through different organs in man in ml. per 100 g. per min obtained by various methods

	Blood Flow ml./100 g./min
Thyroid	560
Kidney	150
Liver	150
Heart (coronary circulation)	100
Intestine	70
Brain	65
Spleen	40
Stomach	25
Hand	7 to 12
Finger	15 to 40
Forearm (mainly muscle)	1 to 3
Leg (mainly muscle)	1 to 2

3. **Circulation Time.**—Various methods are available for estimating the time taken by the blood to pass from one point to another of the circulatory system. The total circulation time may be determined by injecting the dye fluorescein into a vein and timing its appearance in the corresponding vein on the opposite side. This gives the time taken to go round the systemic and pulmonary circulations. Thus it has been found that the jugular to jugular time is about 22 sec, and the time required for the dye to reach the veins in one foot when injected into the other foot is about 28 sec. The total circulation time is shorter in small animals but in all species it corresponds closely to the time of 28 heart beats. Saccharine (20 per cent. solution) and decholin (2 per cent. solution), both of which can be detected as a taste when they reach capillaries in the tongue, are used to determine the vein-heart-tongue capillary time which is normally 11 to 14 sec. Ether or acetone which can be detected by their odour in the expired air may be used to measure the vein to lung time, which is approximately 6 seconds. All these data apply to basal conditions

and must be much reduced in exercise. More recently radioactive material detected by an ionization chamber has been used to determine the circulation time between various points in the circulation such, for example, as the time taken for blood to flow along the veins of the leg from the dorsal venous arch on the foot to the femoral vein at the inguinal ligament.

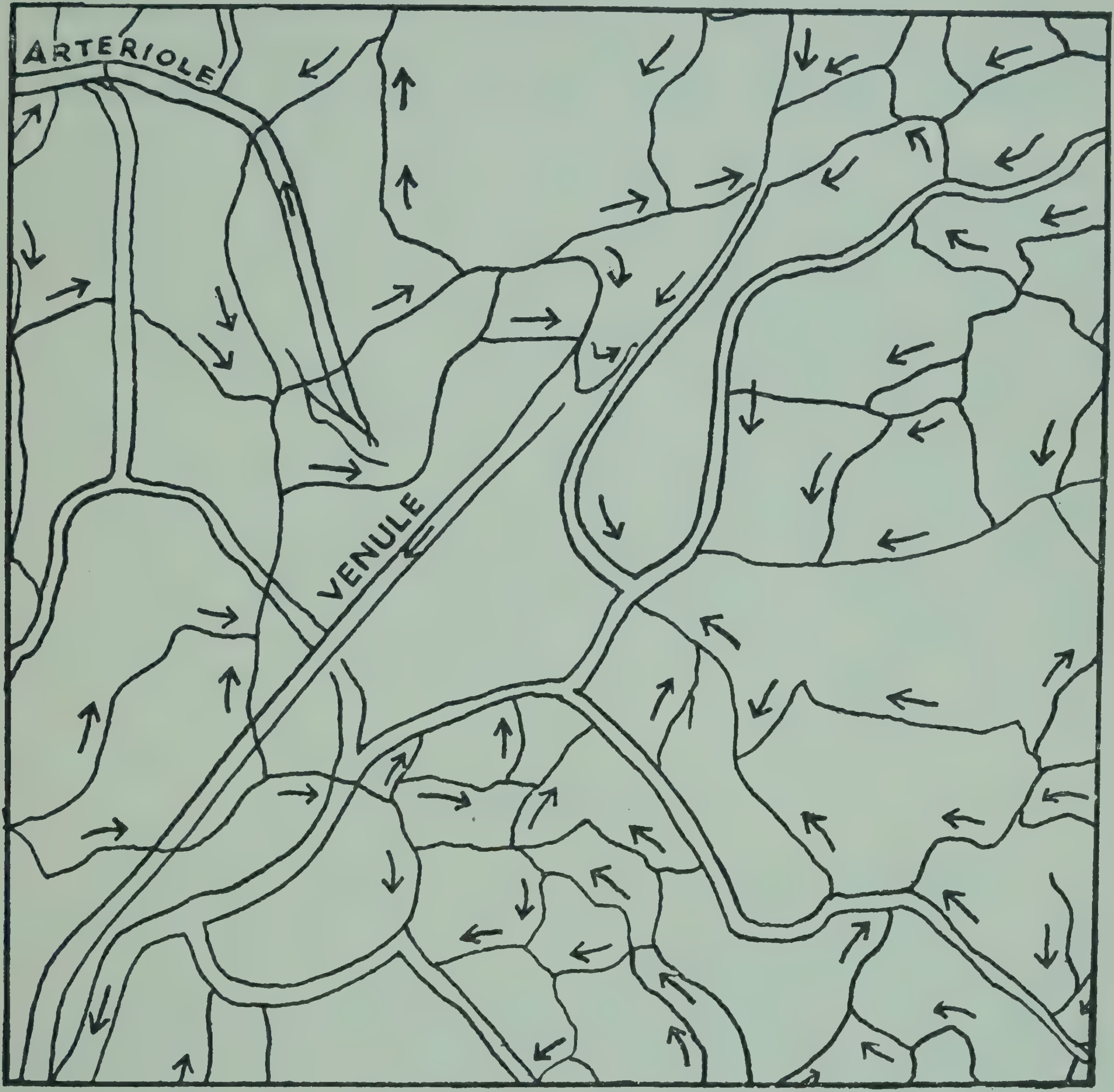


FIG. 29, 28

A drawing of the minute vessels in the web of a frog's foot (× 80), showing the direction of blood flow. The table below indicates the rapidity and direction of flow in a particular capillary observed during 12 min.

Time	*Appearance of Flow*	*Time*	*Appearance of Flow*
12.20	Sluggish ⟵	12.26	V. rapid ⟵
21	Stop.	28	Do.
22	Intermittent.	29	Slower and irregular.
23	V. rapid ⟵	30	
24	Do. ⟵	31	? Pulsatile flow.
25	Do. ⟵	32	Irregular.

Because of the many different routes by which blood can pass between any two points in the circulatory system, the circulation time can at best give but rough information about changes in the circulation, and none at all about how and where these changes have occurred. Although many cases of cardiac failure have, as might be expected, an unduly prolonged circulation time, nevertheless in failure associated with a high cardiac output (p. 446) the circulation time may actually be decreased.

OBSERVATION OF THE FLOW IN THE BLOOD VESSELS

When the frog's tongue, lung, or web of the foot is observed under a microscope a network of vessels is seen (Fig. 29, 28). In the arterioles the red cells move forward in jerks corresponding to the beats of the heart but the pulsatile nature of the flow is lost in the capillaries. If the capillaries are narrow the red cells follow one another in single file and may become elongated, but they recover their shape on reaching a wider vessel. The capillaries show great variations in calibre, with corresponding alterations in flow. In the venules the red blood cells pass along in a steady axial stream, leaving a clear peripheral stream of colourless plasma.

The capillaries in the nail bed of the human finger are easily observed under the microscope if the skin is illuminated by an oblique beam of light. A drop of oil makes the skin more translucent by replacing the air in the surface layers of the epidermis. The capillaries are of the order of 10 μ in diameter but the capillary loops vary greatly in size, in shape and in lumen, the venous side being usually much wider than the arterial side. Spontaneous alterations in the capillary diameter and the rate of flow of corpuscles occur rarely at the bases of the nail, but in the back of the hand they are very frequent, capillaries opening and closing in the absence of any apparent stimulus. Thus although the pulsatile flow characteristic of the arterial system is absent in the capillaries the flow through these vessels is anything but continuous.

REFERENCES

BURTON, A. C. (1954). Relation of structure to function of the tissues of the wall of the blood vessels. *Physiol. Rev.* **34,** 619–642.

DORNHORST, A. C., HOWARD, P. & LEATHART, G. L. (1952). Respiratory variations of blood pressure. *Circulation,* **6,** 553–558.

FRANKLIN, K. J. (1937). *A Monograph on Veins.* Springfield: C. C. Thomas.

HOOKER, D. R. (1921). Evidence of fundamental activity on the part of the capillaries and venules. *Physiol. Rev.* **1,** 112–140.

KROGH, A. (1929). *Anatomy and Physiology of the Capillaries.* New Haven: Yale University Press.

LANDIS, E. M. (1934). Capillary pressure and capillary permeability. *Physiol. Rev.* **14,** 404–426.

LANDIS, E. M. & HORTENSTINE, J. C. (1950). Functional significance of venous blood pressure. *Physiol. Rev.* **30,** 1–32.

LEWIS, T. (1927). *The Blood Vessels of the Human Skin and their Responses.* London: Shaw & Sons.

PICKERING, G. W. (1955). *High Blood Pressure.* London: J. & A. Churchill.

POISEUILLE, J. L. M. (1842). Recherches expérimentales sur le mouvement des liquides dans les tubes de très petits diamètres. *C. R. Acad. Sci. Paris,* **15,** 1167–1186.

STEWART, W. K. (1952–53). The physiological effects of gravity. In *Lectures on the Scientific Basis of Medicine.* Vol. 2. London: Athlone Press.

WIGGERS, C. J. (1952). *Circulatory Dynamics, Physiologic Studies.* New York: Grune & Stratton.

CHAPTER 30

VASOMOTOR CONTROL

NERVOUS REGULATION OF THE BLOOD VESSELS

THE existence of a nervous control of the blood vessels was first shown in 1852 by the French physiologist Claude Bernard who cut the cervical sympathetic in the neck of a rabbit and found that the blood vessels in the ear of the same side became dilated and the ear warmer. Since electrical stimulation of the peripheral end of this nerve, that is the end going to the ear, caused a constriction of the vessels and a lowering of the temperature of the ear the blood vessels must normally have been kept constricted by impulses continually passing along the nerve.

Vasomotor Centre.—The central origin of the vasoconstrictor nerve impulses can be traced by making sections through the brain stem at various levels, starting at the cranial end and gradually passing caudad. There is no alteration of blood pressure if the cut passes above (cranial to) the fourth ventricle, but if the lower (caudal) part of this area is damaged there is a widespread vasodilatation and a fall of blood pressure. Destruction of a small region near the facial colliculi, half-way down the floor of the fourth ventricle (see p. 847), causes a large and immediate fall of blood pressure. Stimulation of the facial colliculus, or of the grey matter near it, gives a rise of blood pressure. Stimulation of the grey matter at the inferior angle of the fourth ventricle causes a fall of pressure but this is not satisfactory proof that there is a vasodilator centre. Such experiments show that in this region of the medulla there is a large number of cells and fibres concerned with the transmission of impulses to blood vessels and this area is, therefore, called the *vasomotor centre*. Afferent impulses are constantly reaching the vasomotor centre from all parts of the body, especially from the pressor receptors in the carotid sinus and aortic arch and from the cardiac and respiratory centres in the medulla itself; efferent impulses are thus initiated or modified and appropriate adjustments made in the circulatory system.

When the cord is cut through in the cervical region the outflow of vasomotor impulses is stopped and there is a profound fall of blood pressure, but if the animal can be kept alive for several days the pressure rises again. It is usually assumed that this rise is produced by the activity of the spinal vasomotor centres in the lateral columns of grey matter. These spinal centres are normally subordinate to the main vasomotor centre, being controlled by fibres which pass down from the medulla in the lateral columns of the cord.

Electrical stimulation of the lateral and posterior parts of the hypothalamus in animals, and of the premotor cerebral cortex, produce vasoconstriction and a rise of blood pressure. In man, mental work as well as all kinds of emotional and psychic disturbances are accompanied by alterations of the circulation, the pallor of anger, for example, and the blush of shame, that seem to indicate connexions between the higher and lower centres.

When a record of the blood pressure of an animal is made as described on p. 483, or in man when a photographic record is obtained from an electro-manometer connected directly with an artery, the trace usually shows fluctuations in time with the respiratory movements (see first part of Fig. 30, 5, and Fig. 31, 2). These Traube-Hering waves are usually said to be due to irradiation from the respiratory centre to the vasomotor centre though they may be the result of variation in activity of a centre in the mid-brain. Occasionally fluctuations of much lower frequency (Mayer waves) occur when the animal is anoxic or has a low blood pressure. The Mayer waves disappear if the carotid bodies are denervated (Fig. 30, 1) or put out of action by injection of acetic acid and they are accounted for

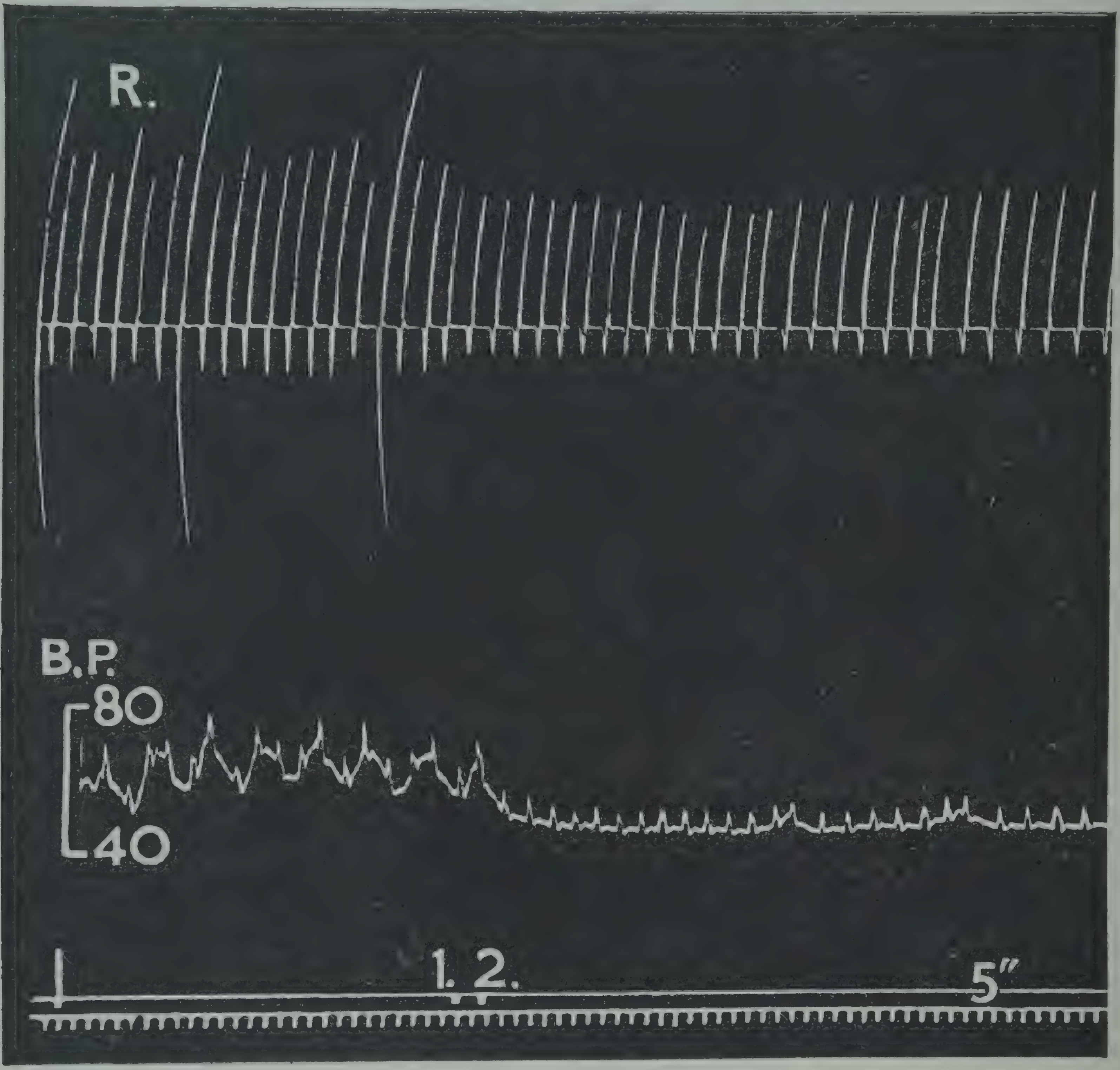


Fig. 30, 1

Cat 2·9 kg. Bilateral vagotomy. Previously bled 65 ml. Records from above downwards:—Respiration; arterial blood pressure; signal line; time in 5 second intervals. The first part of the blood pressure trace shows Mayer waves each about 20 sec; the interval between respirations on the upper trace is 7·5 sec. At 1 and 2 the carotid sinus nerves were sectioned. Note fall in blood pressure and the disappearance of the Mayer waves following this procedure. (B. Andersson, R. A. Kenney & E. Neil (1950), *Acta physiol. scand.* **20,** 206).

as follows. When the blood pressure is low the chemoreceptors in the aorta and carotid bodies are excited (p. 513) and impulses travel in the vagus and carotid sinus nerves to the vasomotor and respiratory centres. As a result the blood pressure rises and respiration improves; the chemoreceptor drive is thus reduced, the blood pressure falls, respiration diminishes and the whole cycle is repeated.

VASOMOTOR NERVES

Vasoconstrictor Nerves.—The pathways by which the vasomotor impulses leave the central nervous system can be shown by making sections through the spinal cord from the lower (caudal) end upwards. Sections below the second lumbar segment have no effect on blood pressure but section higher up is followed by a fall of blood pressure due to vasodilatation in those parts of the body innervated from segments below the level of the section. The higher up the cord the section is made the greater is the fall of blood pressure, the effect being maximal when the cut is made through the first thoracic segment. The vasoconstrictor impulses must, therefore, pass down from the medulla into the cord and leave by the thoracic and upper lumbar nerves. Stimulation of these nerves causes vasoconstriction and a rise of blood pressure. The path of the impulses can be traced further because electrical stimulation of the anterior spinal roots of the thoracic and upper lumbar part of the cord, or of the corresponding white rami communicantes, or of any of the sympathetic nerves arising from them, causes a vasoconstriction in the area supplied. Stimulation of the peripheral ends of the cervical or lower lumbar or sacral nerves has no effect on blood pressure.

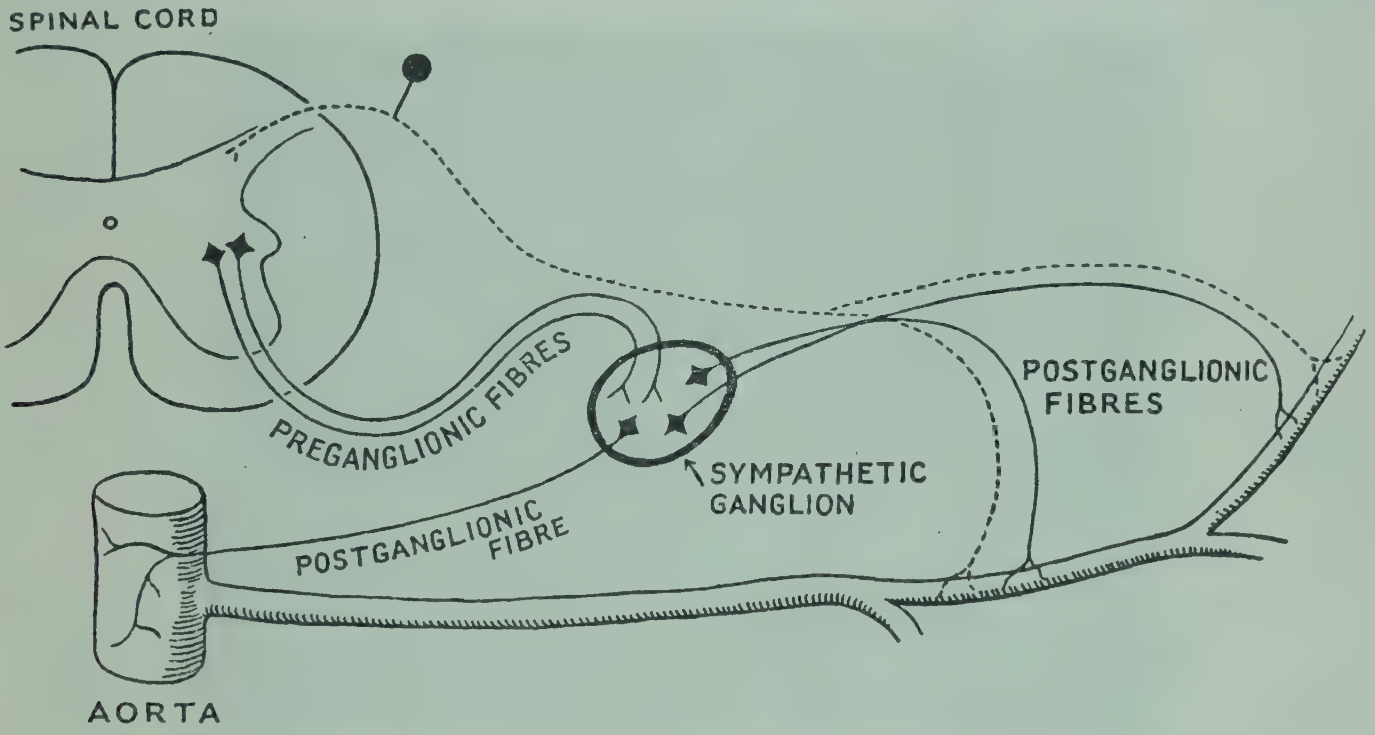


FIG. 30, 2

Sympathetic nerve supply to the blood vessels. (*After* Cowdry.)

The position of the synapses on the vasoconstrictor pathway was investigated by Langley by painting the sympathetic ganglia with nicotine which blocks the transmission of the nerve impulse across the synapses. The fibres leave the lateral horn of the grey matter in the anterior (ventral) root, passing by way of the white ramus communicans to the sympathetic ganglion (*preganglionic fibres*) (Fig. 30, 2). Here new fibres (*postganglionic* *fibres)* arise, some going straight to the main blood vessels, others returning to the mixed spinal nerves, to be carried in them to the blood vessels in the periphery. Although the blood vessels anastomose across the mid-line these vasoconstrictor fibres are confined strictly to their own side of the body.

The abdominal organs are richly supplied with sympathetic fibres through the splanchnic nerves. The fibres arise from the lower thoracic part of the cord, passing through the sympathetic chain to the cœliac ganglion and to the inferior mesenteric ganglia. From here postganglionic fibres reach the abdominal organs along the blood vessels. Denervation of a single organ in the body does not affect the general mean arterial pressure because the resulting vasodilatation produces only a small alteration of the circulation, for which reflex adjustments can easily compensate. Section of the splanchnic nerves, however, is followed by an immediate and profound fall of pressure, not usually permanent. Electrical stimulation of the peripheral ends of the splanchnic nerves (Fig. 30, 3) causes a marked vasoconstriction, which is shown by an increase in arterial pressure and a fall in volume in many of the abdominal viscera. These observations indicate the importance of the splanchnic area in the maintenance of arterial pressure.

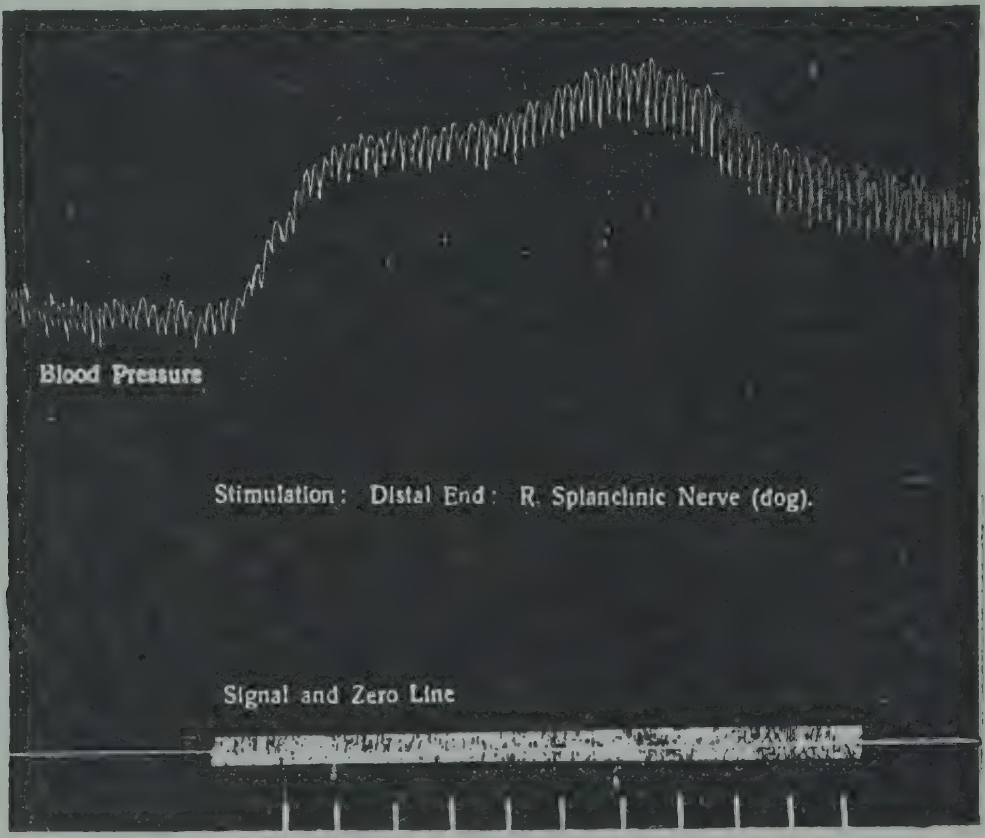


FIG. 30, 3

The effect of stimulation of the peripheral end of the right splanchnic nerve. The time is marked in seconds and ten seconds at the lower edge of the tracing.

(*By courtesy of E. W. H. Cruickshank.*)

Vasodilator Nerves.—Vasodilatation can be brought about, as in Claude Bernard's original experiment, by a reduction in the outflow of vasoconstrictor impulses, that is by a reduction of vasomotor tone. There are however, true vasodilator nerves in the parasympathetic system and possibly also in the sympathetic system though they have little to do with the regulation of arterial pressure; their effects are normally overshadowed by the preponderance of vasoconstrictor impulses. Stimulation of the following parasympathetic nerves produces a vasodilatation in the part supplied: chorda tympani to the submandibular gland, small petrosal nerve to the parotid gland, lingual nerve to the tongue and the pelvic nerves to the external genitalia. Stimulation of the sympathetic nerves to the heart causes an increased heart rate and a dilatation of the coronary arteries (p. 457). This does not, however, prove that the coronary arteries are supplied by sympathetic vasodilator nerves and it is difficult to demonstrate sympathetic vasodilator fibres in other regions. However, if the vasoconstrictor effects are first abolished by ergotoxine, stimulation of the splanchnic nerves may produce a fall of blood pressure. It has been clearly shown that the pronounced vasodilatation in the muscles of the human forearm that occurred in a faint (p. 523) is mediated by sympathetic vasodilator impulses.

In animals stimulation of the hypothalamus and of certain parts of the cerebral cortex is followed by vasodilatation confined to skeletal muscle and abolished by sympathectomy or small amounts of atropine. Although such impulses could account for some of the alterations in the circulation that occur even before exercise has begun their role in physiology is unknown.

ANTIDROMIC IMPULSES AND VASODILATATION

It has been known for many years that stimulation, either electrical or mechanical, of the peripheral end of the divided posterior root of a spinal nerve in an animal produces dilatation of the vessels in the corresponding areas. More recently cutaneous vasodilatation has been demonstrated following stimulation of posterior nerve roots in man. It is now known that such impulses do not travel in efferent fibres (that is fibres originating in cells lying in the cord itself), for the phenomenon still occurs even when the posterior nerve root has been divided between the ganglion and the spinal cord and after time has been allowed for efferent fibres to degenerate. When the posterior root is divided distal to the ganglion and the fibres are allowed to degenerate, stimulation of the peripheral end of the cut nerve no longer produces dilatation at the periphery. In other words the cell stations for the fibres responsible for vasodilatation are in the posterior root ganglia and these fibres are indistinguishable from ordinary afferent fibres. Since the impulses that produce dilatation are evidently passing in afferent nerves in a direction opposite to that taken by sensory impulses they are said to be *antidromic*.

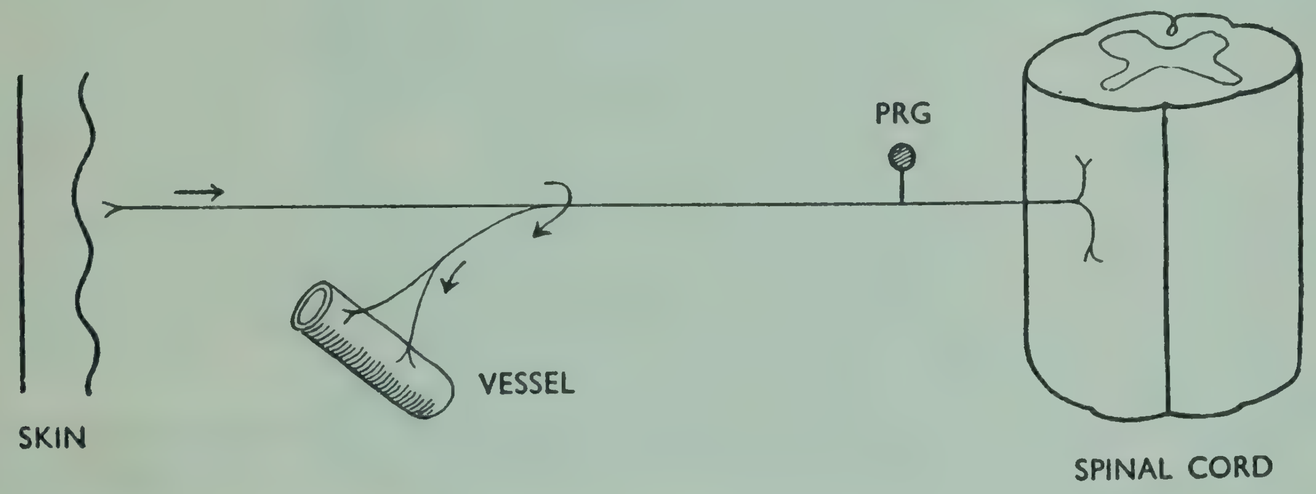


FIG. 30, 4

Diagram to illustrate the probable pathway (shown by the arrows) of impulses set up when an irritant is placed on the skin. PRG, posterior root ganglion.

It is very doubtful if antidromic impulses arising in spinal ganglia play any part in regulating circulatory changes in peripheral tissues, but the cutaneous vasodilatation that follows, for example, the application of an irritant like mustard to the skin is probably brought about by an *axon reflex* occurring in afferent (sensory) nerve fibres as indicated in Fig. 30, 4. Such dilatation no longer occurs if the afferent fibres have degenerated following section either of the peripheral nerves or of the posterior roots distal to the spinal ganglia.

Antidromic impulses afford an explanation of the vasodilatation in herpes zoster (shingles) which is associated with an inflammatory disturbance of one or more posterior root ganglia. In this condition both the reddening of the skin with formation of vesicles and the pain are confined to the distribution of one or more sensory nerves.

VASOMOTOR REFLEXES

Some of the methods available for the study of vasomotor changes in man and animals have already been mentioned. Vasodilatation may be indicated by (1) a reddening of the part, (2) an increase in diameter of the small vessels observed microscopically, (3) an increase in the amount of blood flowing through the organ (see p. 501), (4) an increase in the volume of the part, (5) a rise of the local temperature, and (6) an increase in the amount of heat eliminated per unit time. The last two phenomena are particularly noticeable in the hands and feet, normally relatively cool but well supplied with vascular channels (p. 207). Since some of these changes can, however, be brought about by a rise in blood pressure, it is important to record the blood pressure in all experiments in which vasodilatation or vasoconstriction is being studied.

Vasodilator Reflexes.—Vasodilator and depressor reflexes have already been discussed in connexion with the reflex regulation of the heart. In addition to producing a reflex slowing of the heart, a rise of pressure in the aorta or carotid sinus leads to vasodilatation in skeletal muscle and possibly also in the splanchnic area. There are also pressure-sensitive areas in the arteries of the mesentery similar to the carotid sinus, but the reflex effects they originate are relatively small. The head of an animal can be supplied with blood from a heart-lung preparation while the remainder of the body is supplied by the animal's own heart. If now the blood pressure in the head is increased there is a fall in pressure in the trunk due to vasodilatation and slowing of the heart. After the carotid sinus reflex had been discovered it was shown that this effect arises in the sinus and aortic arch, and is not due to increased intracranial pressure as was originally supposed.

In certain individuals the carotid sinus reflex is unusually easily elicited so that slight pressure on the neck, from a tight collar for example, may be sufficient to cause such a fall of arterial pressure that the circulation through the brain becomes temporarily inadequate and the patient faints (Fig. 30, 7).

Vasoconstrictor Reflexes. — Vasoconstrictor or pressor reflexes occur when practically any afferent nerve is stimulated (Fig. 30, 5). Carmichael has investigated the nervous pathways of some of these reflexes by recording

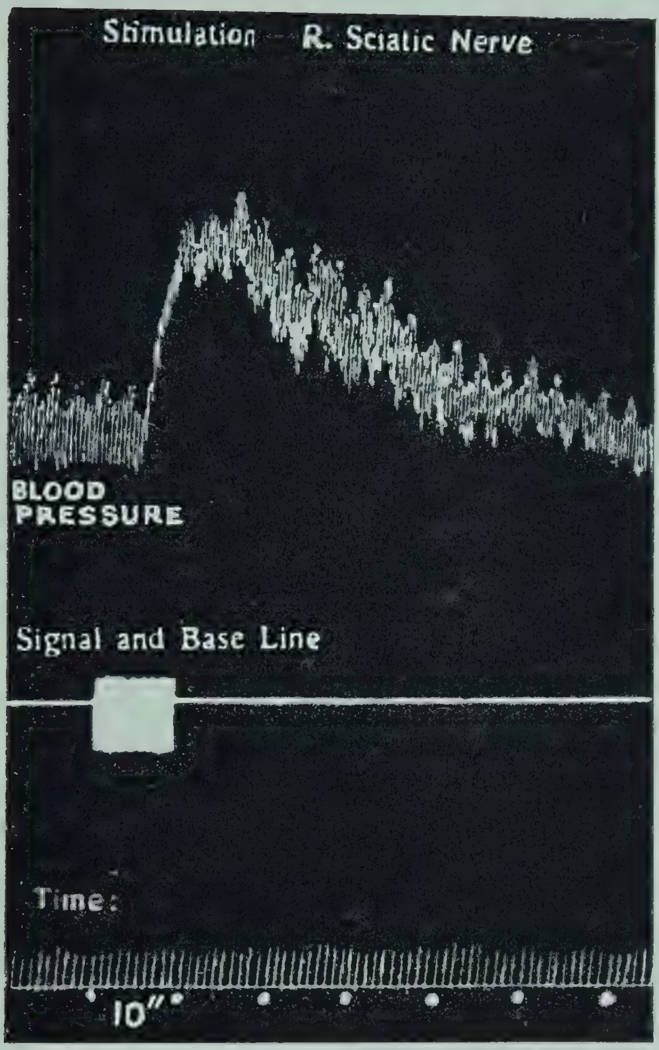


FIG. 30, 5

The effect of stimulation of central end of the sciatic nerve of the dog.
(*By courtesy of E. W. H. Cruickshank.*)

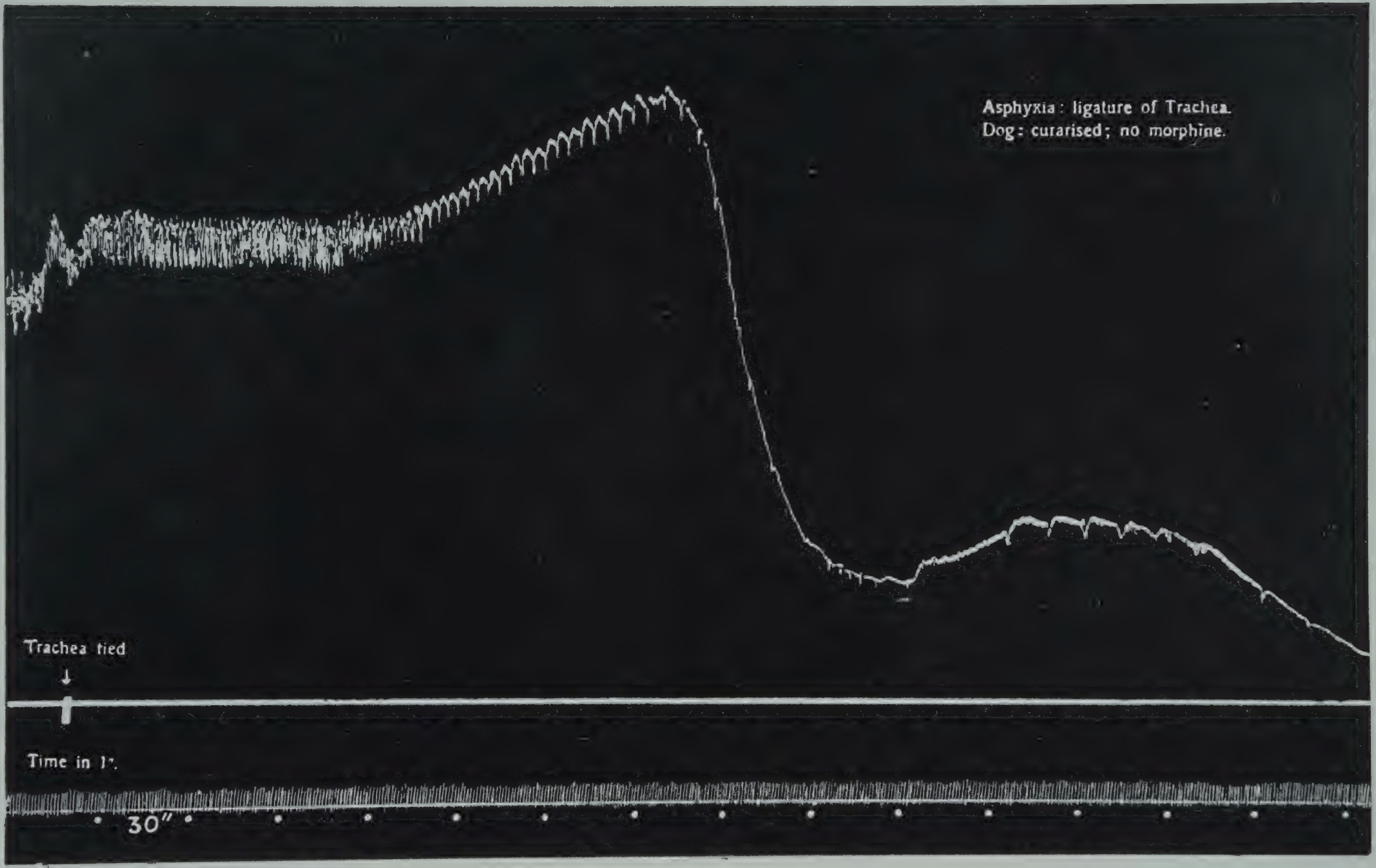


FIG. 30, 6

The effect of asphyxia on the blood pressure of a dog. The animal was anæsthetized with ether before the curare was administered. (*By courtesy of E. W. H. Cruickshank.*)

the volume changes of the toes or fingers in a plethysmograph in normal persons and in patients with various nerve lesions. Any painful stimulus or loud noise, or even the taking of a deep breath, causes reflex vasoconstriction in the skin (Fig. 30, 8). The reflex is abolished if all the nerves to the limbs are sectioned or if the sympathetic nerves alone are divided. This shows that the efferent pathway in man is the same as in animals.

The most important vasoconstrictor reflexes are those originating in the aorta and in the carotid sinuses. A fall of pressure produces reflexly cardiac acceleration and vasoconstriction. The presence of depressor fibres in the vagus is very easily demonstrated in animals, but it was shown by McDowall that this nerve also contains some pressor fibres. When the blood pressure is low after a severe hæmorrhage section of the vagi in the neck produces a further fall of blood pressure. Kenney and Neil have shown that this effect is due to the aortic and carotid chemoreceptors and is absent if these are inactivated by intra-aortic injection of acetic acid. When an animal is bled till its blood pressure is below the threshold (about 60 mm Hg) of the baroceptors in the aorta and carotid sinuses, the chemoceptors in the aorta and carotid bodies discharge and excite the vasomotor centre to maintain vasoconstriction. If the centre is deprived of these impulses by section of the vagi or of the nerves from the carotid bodies the blood pressure falls (see Fig. 30, 1). The discharge from the chemoceptors is probably brought about by anoxia since it diminishes if the animal is made to breathe pure oxygen; the effect of vagal section is also smaller in these circumstances.

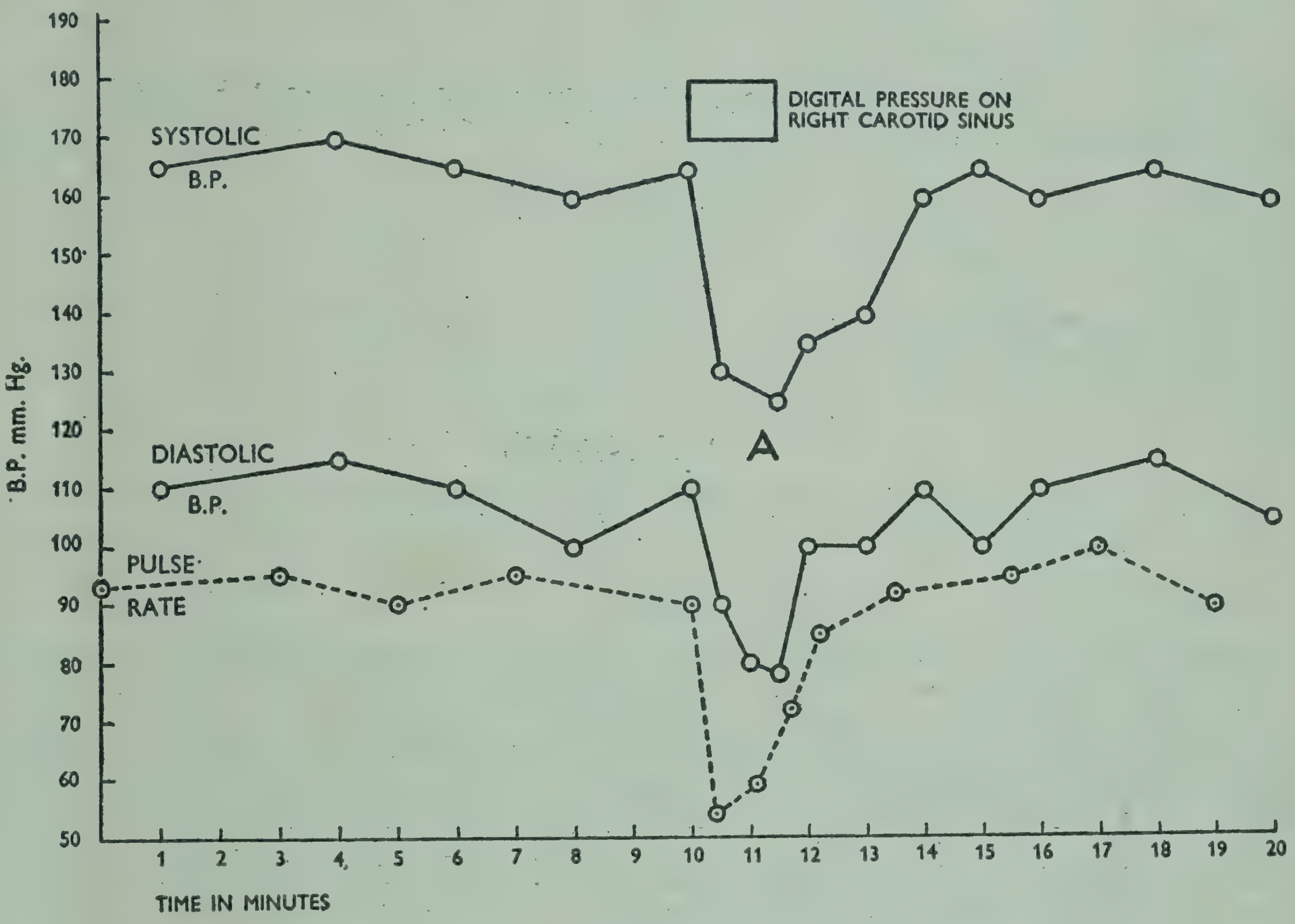


FIG. 30, 7

The behaviour of blood pressure and pulse rate in response to light digital compression over the right carotid sinus for 1½ minutes in a woman aged 34 with high blood pressure. At A the patient complained of feeling giddy and sick.

Lovén Reflex.—Most of the reflex adjustments of the calibre of the blood vessels in response to some local need affect the general blood pressure very little. This is not necessarily because they are small but rather because a dilatation in one part is compensated by a vasoconstriction in other parts of the body. If the volumes of the hind-limbs of an animal are recorded separately while the central end of one femoral nerve is stimulated electrically it will be found that there is a vasodilatation in the ipsilateral limb and a vasoconstriction in the contralateral limb, with usually an increase in the general arterial pressure. This is an example of the type of reflex described by Lovén, the effect of which is to allow a greatly increased blood flow through a part without causing much disturbance of the general blood pressure. It is uncertain to what extent this kind of reflex contributes to the adjustment of the blood flow through an active organ and it has in any case never been demonstrated in man.

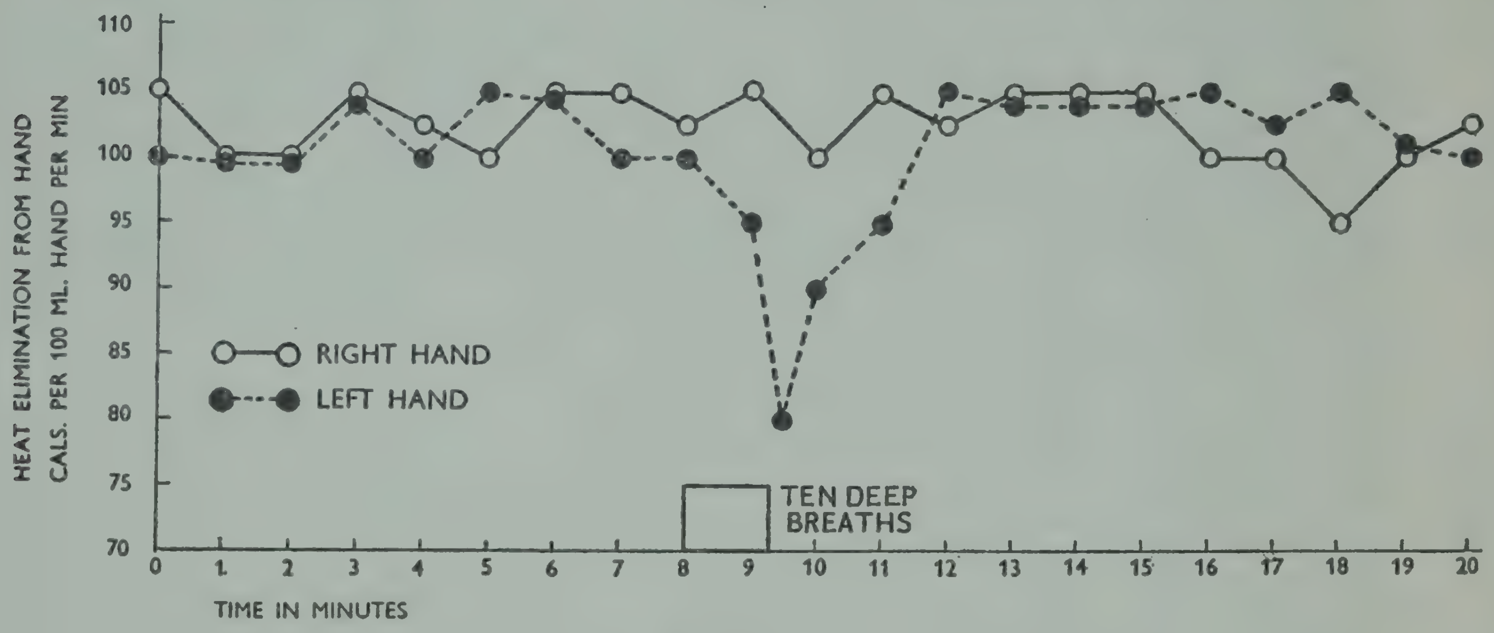


Fig. 30, 8

Calorimetric records of the heat elimination from the right and left hands separately of a healthy man aged 38. Before the record began the median and ulnar nerves on the right side had been infiltrated with procaine. Beginning at the eighth minute as indicated by the open rectangle the subject took ten deep breaths, following which the heat elimination from the left hand fell, indicating vasoconstriction. The absence of vasoconstriction in the right (blocked) hand shows that this is mediated through efferent nerves.

CHEMICAL CONTROL OF VASCULAR TONE

Effect of Carbon Dioxide on the Vasomotor Centre.—The activity of the vasomotor centre is affected, like that of the respiratory centre, by the amount of carbon dioxide in the blood flowing through it. If carbon dioxide is washed out of the blood by over-ventilation then there is a vasodilatation with a fall of blood pressure (Fig. 30, 9). In an anæsthetized animal an excessive quantity of carbon dioxide in the blood passing through the centre causes a rise of blood pressure with a decrease of volume of the muscles and abdominal viscera indicating a vasoconstriction.

Vasoconstriction with a rise of blood pressure is a feature of the early stages of asphyxia produced, for example, by clamping the trachea (Fig. 30, 6). It is followed by a slowing of the heart through the carotid sinus and aortic reflexes, for when these reflexes are abolished by section of the vagi the rise of pressure during asphyxia is much greater. In an anæsthetized animal, however, the initial rise of pressure after clamping the trachea may be small, or absent, but the slowing of the heart and the fall of pressure occur as in the unanæsthetized animal. The reason for this difference is not known. Sometimes in a blood-pressure tracing taken during asphyxia the pressure rises as described and then about half a minute later there is a second rise. Since the latter is abolished by adrenalectomy it is presumably due to an outpouring of adrenaline and noradrenaline.

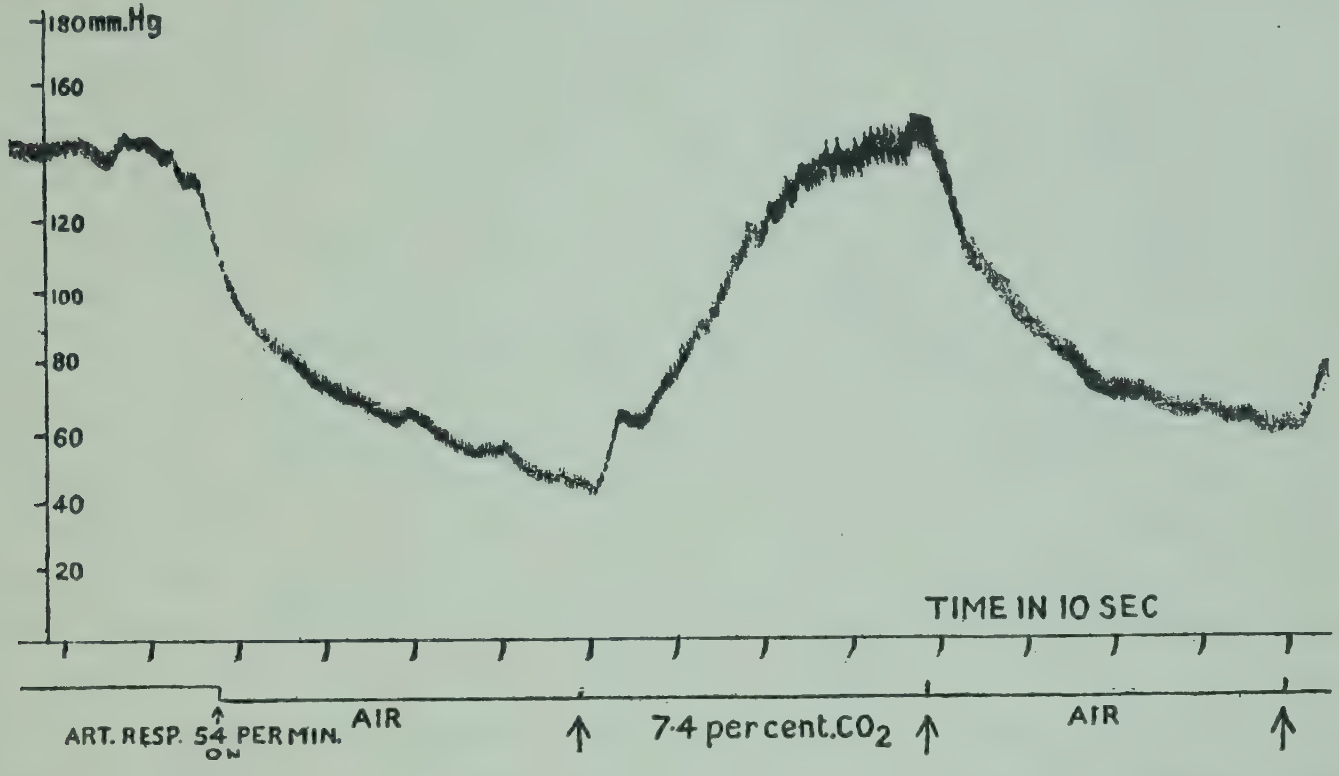


FIG. 30, 9

Blood-pressure record from a decerebrate cat (4·1 kg.). The first part of the trace shows the level of blood pressure, 140 mm Hg, during spontaneous respiration. Artificial respiration with air at the rate of 54 respirations per min by means of a pump produced a marked fall of blood pressure. This is due to hyperventilation with blowing-off of carbon dioxide since the addition of 7·4 per cent. carbon dioxide to the air restores the blood pressure to the original value. Note that hyperventilation in a conscious man does not produce a fall of blood pressure. (*By courtesy of H. P. Gilding.*)

It is obvious that during asphyxia the vasomotor centre suffers from oxygen lack as well as carbon dioxide excess, and both of these produce vasoconstriction. If an animal is made to breathe oxygen containing, say, 7 per cent. of carbon dioxide there is no deficiency of oxygen in the blood yet a rise of blood pressure occurs, together with a reduction in the volume of the viscera.

In the final stages of asphyxia the arterial pressure falls progressively (Fig. 30, 6) presumably because the heart, through lack of oxygen, is unable to maintain its output.

Reactive Hyperæmia.—When the circulation to a limb is arrested by a tourniquet the continuing metabolism of the tissues leads to the accumulation of vasodilator substances that are normally removed in the blood as fast as they are formed. When the blood flow to the limb is re-established by releasing the tourniquet the dilated blood vessels are filled with blood as evidenced by a flush which travels rapidly down the limb to involve the extremities within 2 to 3 seconds. This *reactive hyperæmia* affects also the vessels in the muscles (Fig. 30, 12). It is independent of any nervous connexion and may therefore be used as an index of the capacity of the blood vessels to dilate. When the blood vessels are narrowed by disease and incapable of dilating the flush is delayed or absent. The tourniquet is now generally condemned as a first-aid procedure for the arrest of bleeding, partly because the reactive hyperæmia tends to increase hæmorrhage when the tourniquet is loosened, and partly because prolonged ischæmia does great damage to peripheral nerves.

The vasodilatation of exercise (Fig. 29, 27) is at least partly due to a rise in the local concentration of metabolites resulting from their increased production in the muscles. It cannot be attributed to an increase in hydrogen ion concentration because the pH change likely to occur in the tissues has little effect on the arterioles. Histamine, which has been isolated from most of the tissues of the body, and carbon dioxide, lactic acid, acetylcholine and adenylic acid produced in the course of muscle metabolism (p. 774) are all vasodilator substances and a substance with the physiological properties of histamine has actually been detected in venous blood taken from an arm in which the circulation had been occluded for thirty minutes. These substances may have a direct effect on the blood vessels in the muscles or they may, when they enter the circulation at the venous end of the capillaries, have an indirect effect on the arterioles supplying the part by stimulation of nerve receptors in the small veins or venous capillaries from which impulses are carried by the perivascular network to the arterioles. It is important to note that these peripheral effects of carbon dioxide are the opposite of those produced centrally; when carbon dioxide is present in excessive amounts in the blood passing to the vasomotor centre a vasoconstriction ensues (see p. 514).

Bier's Spots.—When the circulation is arrested in a forearm previously somewhat congested with venous blood, then within a few minutes a number of irregular patches appear in which the dusky colour of the skin has given place to red or white. The red patches are the result of oxygenated blood from deeper vessels being displaced into the skin by blood flowing into the forearm through nutrient arteries in the humerus communicating with anastomotic vessels round the elbow joints. The white patches (Bier's spots) are the result of localized constrictions of the minute vessels of the skin. This phenomenon, which can occur in areas deprived of their nerve supply, is sometimes seen in severe circulatory failure. It also occurs at death when it is part of a generalized vascular constriction.

Adrenaline and Noradrenaline.—When the splanchnic nerves are stimulated the initial rise of blood pressure is probably due to vasoconstriction in the abdominal organs. This is followed about half a minute later by a second rise due to the liberation of adrenaline and noradrenaline from the suprarenal medulla (Fig. 30, 3). A similar effect has already been described in asphyxia (p. 514). Again, when the pressure in the carotid sinus is reduced the rise of arterial blood pressure which follows is due partly to the reflex activity of the vasoconstrictor nerves and partly to liberation of adrenaline and noradrenaline.

When it is injected intravenously into an animal adrenaline produces an immediate and transient fall of arterial pressure at once succeeded by a slight rise. The increase in blood pressure lasts but a few minutes (Fig. 30, 14). It is accompanied by generalized vasoconstriction though the coronary arteries and the vessels of the muscles and splanchnic area are dilated. The vasoconstrictor effect is brought about by the direct action of adrenaline on the blood vessels because it occurs after destruction of the spinal cord and vasomotor nerves but the rise of pressure after an injection of adrenaline is partly due to the increased cardiac output (p. 463).

Injection of noradrenaline is not followed by an initial fall in arterial pressure and the rise in pressure is greater than that following adrenaline. Noradrenaline is therefore a general vasoconstrictor (Fig. 30, 10), and is used as such in clinical medicine.

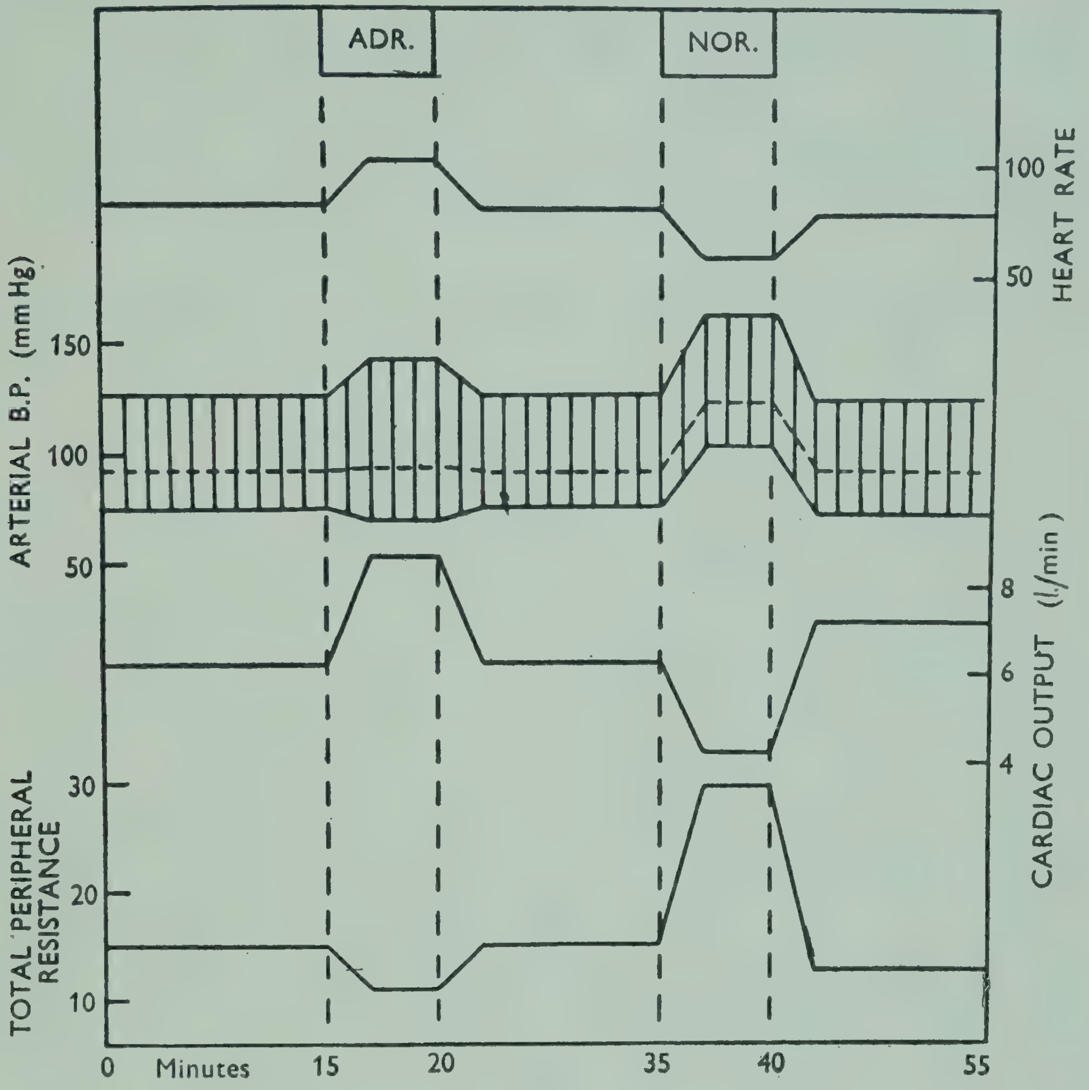


FIG. 30, 10

Diagram showing the effects of intravenous infusions of adrenaline and noradrenaline on heart rate, arterial blood pressure, cardiac output and total peripheral resistance. (The initial transient drop in blood pressure that occurs during infusions of adrenaline is not shown.) Barcroft, H. & Swan, H. J. C. (1953), *Sympathetic Control of Human Blood Vessels*, Monographs of the Physiological Society, Number 1. London: Arnold.

THE BEHAVIOUR OF THE CAPILLARIES

Nervous Control of the Capillaries.—The thin walls of the capillaries consist of a single layer of polygonal endothelial cells. If a capillary stained with methylene blue is examined in the living state the nuclei of the endothelial cells can be seen, and also at intervals along the wall the branched cells of Rouget (Fig. 29, 1), which are regarded by some histologists as modified muscle cells. Most investigators think that the endothelial cells themselves can contract; some say that they can swell up and so occlude the capillary. Histological studies have shown that the capillaries are invested by non-medullated nerve filaments and stimulation of the sympathetic nerve to the rabbit's ear, for example, produces a contraction of the capillaries even when the circulation has been arrested by clamping the main artery. This contraction cannot therefore be a passive effect of a change of blood pressure upon the capillaries. To what extent capillary function is dependent upon the integrity of these nervous connexions is unknown, but experiments have suggested that the capillaries are normally held partially constricted as the result of 'tonic' impulses which reach them in the sympathetic nervous system.

Chemical Control of the Capillaries.—It is clear, however, that the main factors influencing the state of the minute vessels are the contractions and dilatations of the precapillary arterioles and the local concentration of certain chemical substances produced in the tissues either as the result of activity or of injury. The capillaries are nevertheless able to contract independently of the arterioles and venules although how much such alterations in calibre affect the arterial blood pressure is unknown.

Continuous observation of the capillary bed in the mesentery or web of the frog shows that the blood flow through the minute vessels is not a continuous process. In an uninjured preparation areas of the capillary network close down so that blood flow ceases or becomes much reduced while the flow through adjacent areas increases. After a few minutes the capillary bed, previously shut down, reopens and flow is re-established. This periodic redistribution of blood flow between adjacent areas of the capillary bed is sometimes referred to as *vasomotion*. It is believed to be due to variations in tone of the walls of the minute vessels as the result of local changes in the concentration of vasodilator tissue metabolites. During a period of relative ischæmia the concentration of tissue metabolites gradually increases until vasodilatation occurs. The opening up of the capillary bed and the more rapid blood flow serve to remove these dilator substances from the tissue, the tone of the capillary walls is regained, and the blood flow is diminished until the process repeats itself (Fig. 29, 29). There is some doubt as to whether vasomotion occurs in the minute vessels of human skin.

The number of capillaries per unit area of tissue can be approximately determined by examining microscopically sections of the tissues of animals injected before death with India ink. In active muscles the capillaries are not only much more numerous than in resting muscles but also much wider. Thus in the resting muscle of a guinea-pig Krogh found 200 capillaries per sq. mm with an average diameter of $3{\cdot}5\ \mu$, against 2500 sq. mm in an active muscle with an average diameter of $5{\cdot}0\ \mu$. A transverse section of an active skeletal muscle may show three or four times as many capillaries as muscle fibres. Activity is thus accompanied by capillary dilatation and in addition the size of the capillary bed is enormously increased by the opening up of vessels which were closed during the resting state. This opening up of new capillaries is believed to be initiated by the local release of chemical substances.

Contractile Power of Capillaries.—Microscopic examination of the thin-walled capillaries does not suggest that they could contract at all powerfully, but the effectiveness of capillary contraction can be shown in the following way. If the circulation through the arm is completely cut off the injection of extract of the posterior lobe of the pituitary or of adrenaline into the skin of the forearm produces a white area which must, for the reasons given below, be due to capillary contraction. When a sphygmomanometer cuff is placed round the arm and inflated to, say, 40 mm Hg the venous pressure increases quickly to this value and the capillary pressure rises about 10 mm Hg higher (see p. 490). Since an intradermal injection of adrenaline can still produce a white area the capillaries must be capable of contracting against an internal pressure of 50 mm Hg. Experiment shows that they cannot contract against higher pressures. Since normally the pressure at the arteriolar end of a capillary loop rarely exceeds 50 mm Hg (see p. 490), the capillary should be able to act independently of the arteriole in spite of the high pressure at the arteriolar end. If, however, the skin capillaries are first constricted by an injection of adrenaline the venous pressure must be raised to nearly 100 mm Hg before the capillaries give way under the pressure and allow blood to flow back into them from the veins. In other words, once constricted, the capillaries can offer considerable resistance to distension from the venous end, and presumably also from the arteriolar end, of the capillary.

Circulation in the Skin

Skin Colour.—The colour of the skin depends on the amount of blood in the capillaries and in the subpapillary venous plexuses, and also on the state of oxygenation of the blood in these minute vessels. Alterations in the colour of the skin cannot be properly interpreted unless the temperature of the skin is known, for this is governed by the blood flow through it, which depends on the state of the arterioles. Thus a warm, red skin indicates that both arterioles and minute vessels are dilated. When the skin is warm and pale the arterioles are dilated but the minute vessels are constricted, so that the skin is relatively devoid of colour. The raised temperature indicates that the blood flow through the skin is large, while the pallor of the skin shows that the minute vessels are empty. The blood is, in fact, passing rapidly from arterioles to small veins through arterio-venous anastomoses (shunts) (Figs. 30, 11 and 13). When the skin is cold (blood flow reduced) it may be pale, red or blue. A cold, pale skin is evidence of arteriolar constriction and closure of minute vessels. In both cold red and cold blue skins the arterioles are constricted and the minute vessels full. The blue colour is due either to back flow of venous blood into the dilated vessels or to stagnation. Cold red skin occurs only at temperatures at which dissociation of oxyhæmoglobin virtually ceases; it can be readily produced by placing the hand in water at 5° C.

Mechanical Stimulation of Skin Vessels.—Certain special features of the capillary circulation in human skin have been demonstrated by Lewis. If the forearm is stroked gently with a blunt-pointed instrument the skin is flushed for about fifteen seconds and then a white line appears. The latter is caused by capillary constriction, since it occurs in the exact path of the instrument and does not have the irregular outline produced by constriction of arterioles. Further, even when the circulation is first completely occluded by a sudden inflation to 200 mm Hg of a sphygmomanometer cuff placed round the upper arm, stroking of the forearm still produces a white line. If, in these circumstances, the arterioles were to contract they could only expel their contents into the capillaries and so a red line would be produced. The only possible explanation of the blanching produced on stroking the skin is that the capillaries have contracted or have become occluded by swelling of the endothelial cells. Since the blanching occurs even when the nerves of the skin have degenerated it must be a local response to the stimulus.

The Triple Response.—If the skin is stroked very heavily, or scratched, a more striking vascular response occurs, most easily investigated in persons with sensitive skins. This response has three main phases and Lewis called it the *triple response.* (1) After a short latent period a *red line* appears in the track of the stroke. It is due to capillary dilatation and may tend to become blue if the blood stagnates. This is a local response, because it can be obtained after degeneration of the cutaneous nerves.

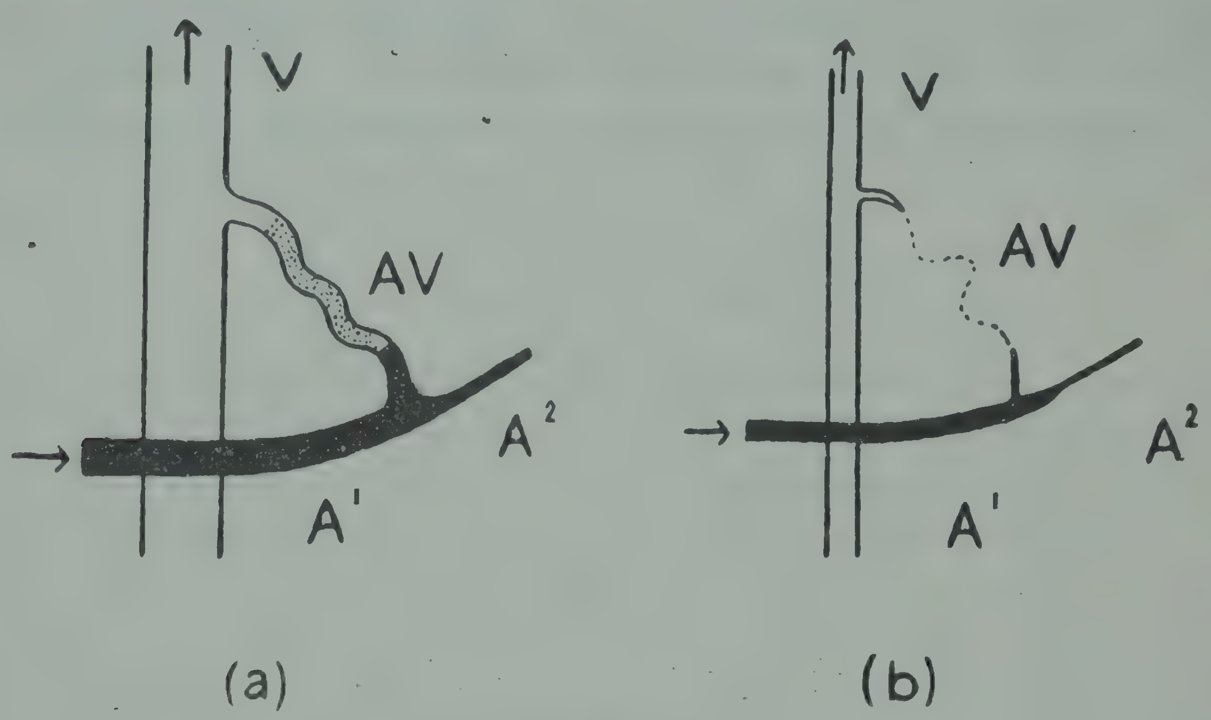


Fig. 30, 11

Reaction of anastomoses and associated vessels in the rabbit's ear to lowering of body temperature. V, vein. AV, anastomosis. A^1, artery. A^2, arteriole. (a) before and (b) after lowering of body temperature (R. T. Grant (1929), Arteriovenous anastomoses, *Heart,* **15** to **16**, 294).

There may be for a short time a white reaction on each side of the red line. (2) Nearly half a minute later there is a flush or *flare* over a considerable area on either side of the red line. This is believed to be due to arteriolar dilatation for the following reasons. The area of the flare is mottled and has an irregular edge, as might be expected from the anatomical distribution of the arterioles. Since the flare persists for some time, and the skin temperature rises locally, the blood must be flowing quickly through the skin. The flare occurs even if the nerve trunk supplying the area of skin has been blocked by an anæsthetic but not if the local nerve fibres have degenerated. It is likely that it depends on axon reflexes. (3) A little later a weal, that is a raised œdematous area of skin, develops along the red line, reaching its maximum size in about five minutes. As the weal becomes larger the vessels are compressed by the œdema fluid and the area becomes pale. This weal is due to the escape of fluid into the tissue owing to the increased permeability of the minute vessels because (*a*) the fluid escaping into the tissues has a protein content approaching that of blood plasma, and (*b*) the dye trypan blue given intravenously escapes into the weal but not into the tissues elsewhere.

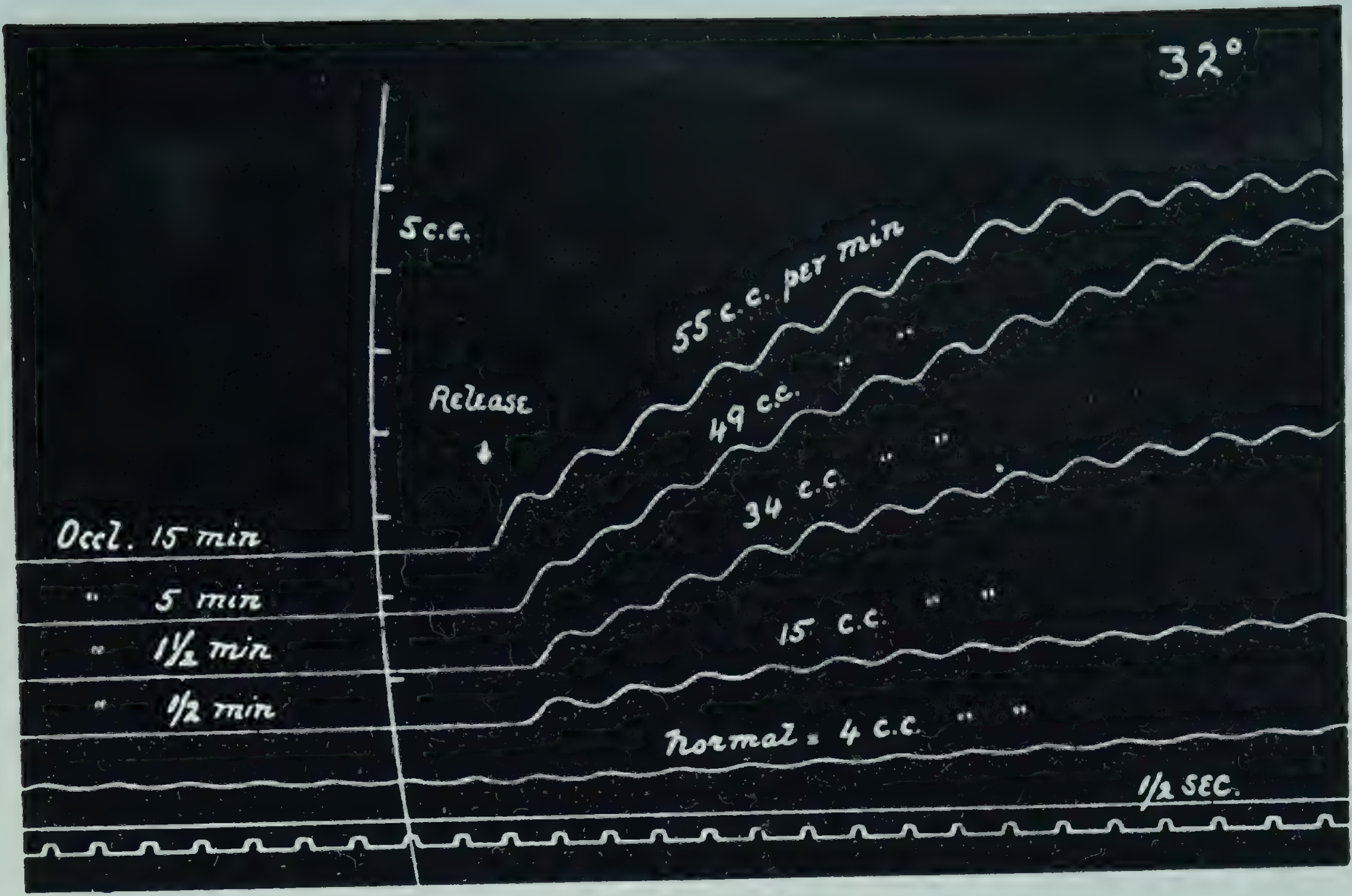


FIG. 30, 12

Blood flow through the forearm determined by a plethysmograph before (normal) and at various times after the temporary arrest of the circulation by raising the pressure in a cuff round the upper arm to 200 mm Hg for ½, 1½, 5 and 15 min. The occluding pressure was released at the arrow and the resulting increase in blood flow immediately determined by measuring the increase in volume of the forearm following obstruction of its venous outflow. The reactive hyperæmia is proportional to the time of circulatory arrest. For details of method see Fig. 29, 26 (T. Lewis & R. Grant (1925–26), Reaction of hyperæmia, *Heart*, **12**, 86).

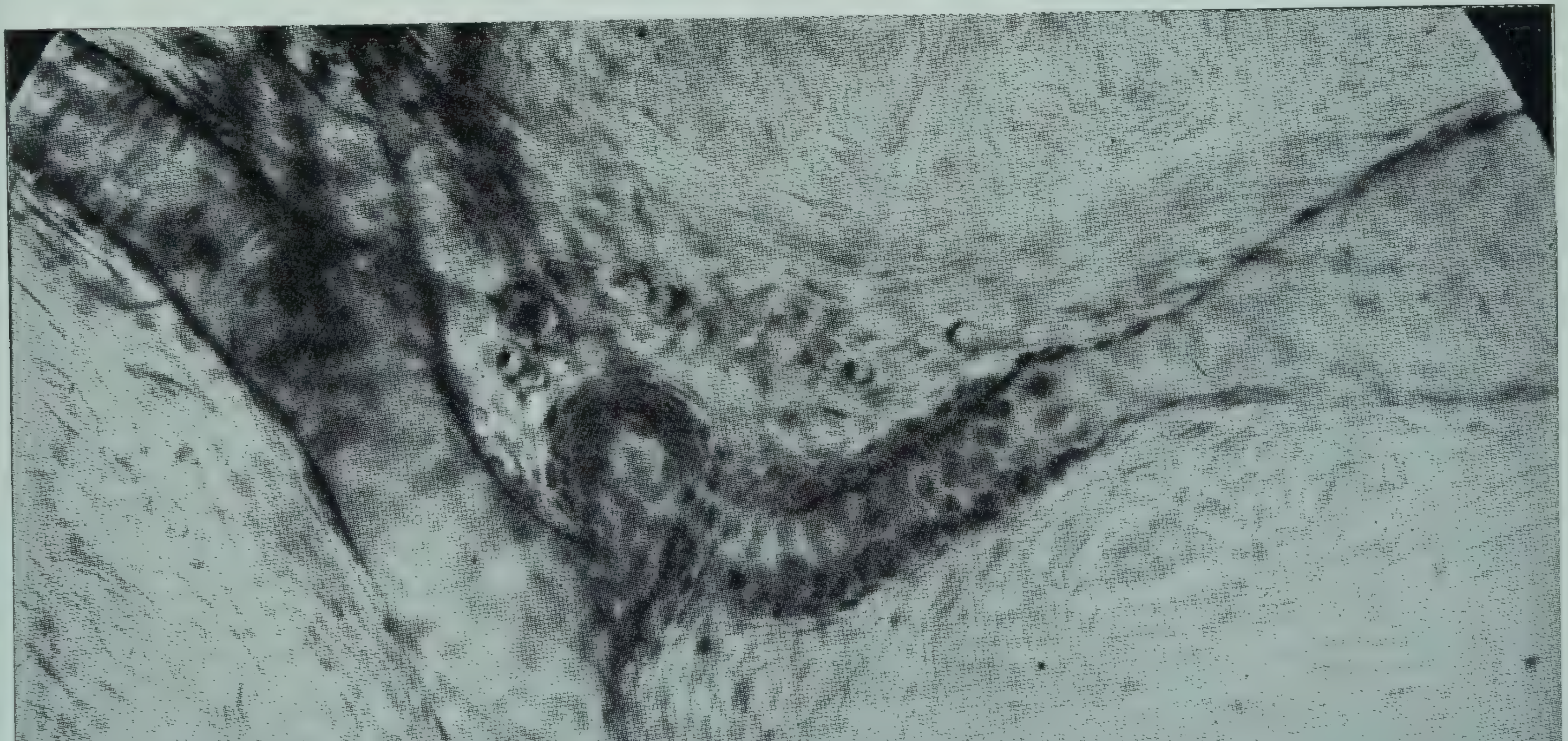


FIG. 30, 13

Untouched photomicrograph (× 300) of an arteriovenous anastomosis in the sole of the foot (from a section 200 μ thick). To the left of the figure an artery runs upwards, the nuclei of its circular muscle fibres are in focus at the top of the figure. It gives rise to an anastomosis that coils on itself and passes to the right to widen out into a thin-walled vein. Note the numerous rounded nuclei in the thick wall of the anastomosis (R. T. Grant (1929), Arteriovenous anastomoses, *Heart*, **15** to **16**, 409).

[*Facing page* 520

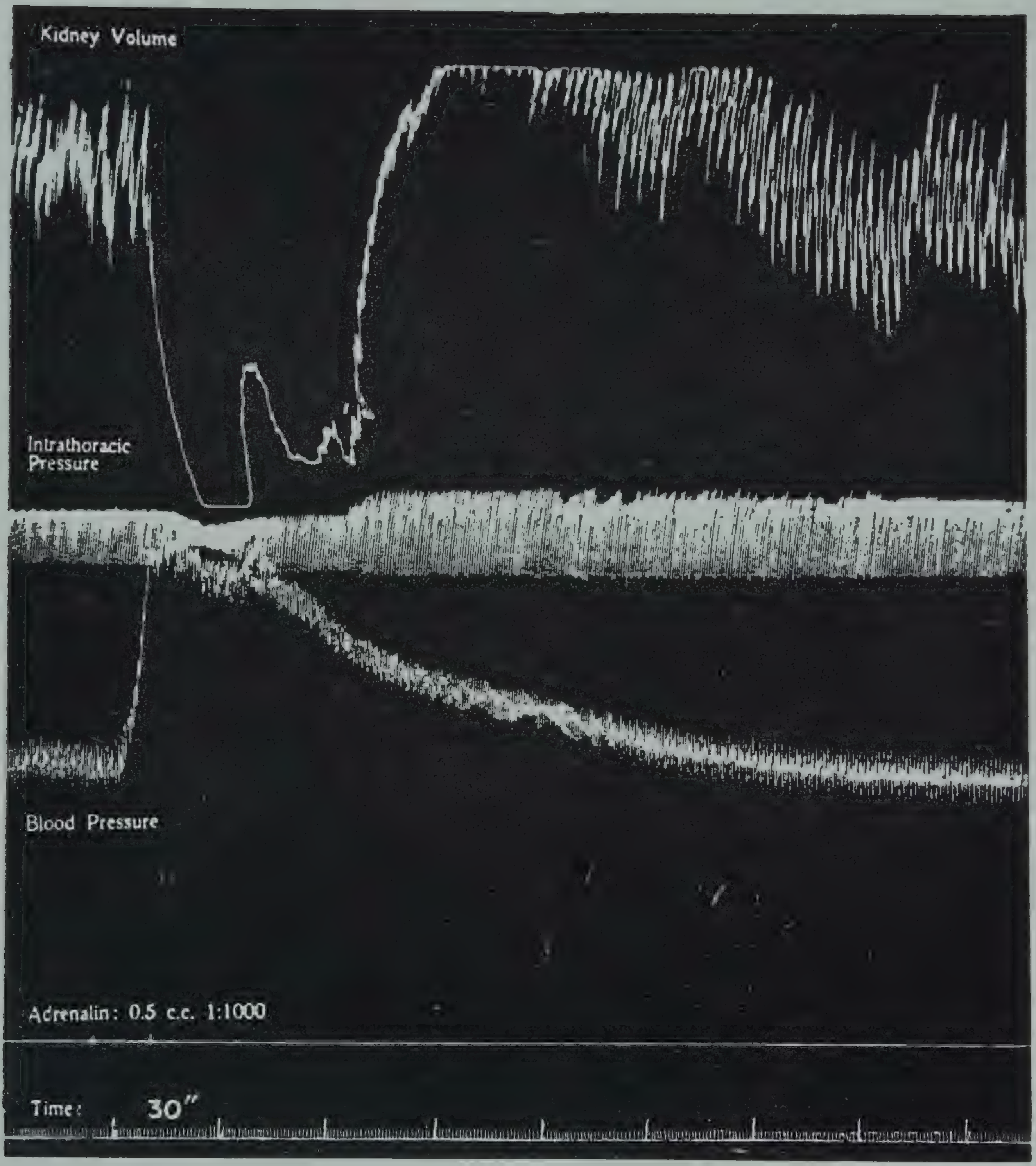


FIG. 30, 14

The effect of adrenaline on kidney volume, intrathoracic pressure and blood pressure of the dog.

(*By courtesy of E. W. H. Cruickshank.*)

Lewis's H-substance.—Lewis has shown that a triple response—red line, flare and weal—can be obtained by injection of minute amounts of histamine into the skin. Further, by stroking the skin firmly at one place and injecting histamine simultaneously at another it was found that the times of appearance and disappearance of the responses in both places were similar. This is circumstantial evidence that the triple response is caused by histamine produced or released as the result of local damage of the skin. Histamine has actually been detected in extracts of human epidermis. A large number of irritants, ranging from insect bites to acids and alkalies, and freezing or burning produce the triple response. This can best be explained by supposing that in all these circumstances damage to the skin liberates the same chemical substance ('H'-substance).

If the circulation is arrested and an injection of histamine is made into the skin of the forearm and then, after ten minutes, the tourniquet is released, the reactive hyperæmia is followed by the appearance of a histamine flare. The times at which this flare appears and disappears are roughly the same whether the injection is made one or ten minutes before the release of the tourniquet. The flare produced by stroking during circulatory arrest is similarly independent of the time elapsing before restoring the circulation. This suggests that the flare is not produced by the mechanical stimulus itself but by a chemical substance which is liberated from the tissue while the circulation is arrested, but produces its effects on the local nerve endings only after the circulation is restored.

Certain individuals whose skin is abnormally sensitive to an injection of histamine and in whom 'H'-substance is liberated in the skin with unusual ease develop all the features of the triple response as the result of comparatively mild trauma. If the reaction is particularly severe such a person complains of headache, and the wealing is associated with a flushing of the face and secretion of acid by the stomach—all effects which are produced by the subcutaneous injection of histamine.

When the arm is plunged into hot water the arterioles and capillaries dilate, partly owing to the direct action of heat on them and partly to liberation of 'H'-substance. If the hand is placed in very cold water, say 5° C, then it becomes bright red, the arterioles are constricted and the capillaries are dilated. Cyanosis does not occur because oxyhæmoglobin is not easily reduced at low temperatures, and in any case the metabolic activity of the skin is greatly depressed.

Effects of Cold upon the Skin.—When the skin is exposed to cold vasoconstriction occurs, blood flow is reduced and heat loss is diminished. The vasoconstriction is brought about in three ways. There is first of all a direct response of the superficial vessels to cold ('direct vasoconstriction'); secondly a generalized, though transient, vasoconstriction for which both afferent and efferent pathways are in the nervous system ('reflex vasoconstriction'); and thirdly a more persistent generalized vasoconstriction which is brought about by an increase of sympathetic tone initiated by cool blood from the exposed part flowing through the brain ('indirect vasoconstriction'). Although the effect of these reactions is to conserve body heat the temperature of the skin may continue to fall. This leads to further vasoconstriction and greater fall of temperature, with the result that the temperature of the part, falls closer and closer to that of its surroundings.

When the temperature of the skin falls below 10° C (50° F) sensation becomes numbed and finally lost. Muscle power in a limb, for example, is reduced and the member becomes completely anæsthetic and powerless so that surgical operations may be performed upon it.

If a finger is kept immersed in water at 15° C (59° F) or less the initial vasoconstriction is succeeded within ten to twenty minutes by a local dilatation, the effect of which is to raise the temperature of the skin by several degrees Centigrade. This vasodilatation is, however, transient, being succeeded by vasoconstriction and a renewed fall of temperature. The sequence is then repeated so that the temperature of the skin rises and falls periodically; the effect, however, of this 'hunting reaction' is that the mean temperature of the part is raised. Since the reaction can occur after complete degeneration of all nerves presumably it cannot result from axon reflexes.

The exposure of part of the body to temperatures of 5° C (41° F) produces pain in the part, redness (cold vasodilatation), swelling and, later, numbness. Blistering of the skin may follow. These are the signs of the 'H'-reaction and indicate that the tissues have been damaged. Very low temperatures cause ice-crystals to form in the tissues ('frost-bite') and to the effects of cold is added mechanical injury.

Effect of Light on the Skin.—Exposure of the skin to light rays containing no ultraviolet radiation produces a reddening or *erythema* of the skin in a quarter of an hour. This is really the effect of heat and the redness, due, of course, to a marked vasodilatation of the skin vessels, passes off in an hour or two. When this heat erythema occurs repeatedly, as it does in the legs of persons who spend much time sitting in front of the fire, it leads to a dark brown pigmentation of the skin in a reticular pattern, which seems to follow the arrangement of veins in the superficial venous plexus (erythema ab igne).

If the skin is exposed to radiation shorter than $0{\cdot}33\mu$ erythema appears after several hours, and persists for about a week. As the redness gradually fades pigmentation and desquamation develop. If the dose of radiation is excessive, œdema and blistering occur. Radiation of longer wavelength, although unable by itself to produce pigmentation, may produce a brownish coloration by the oxidation of pigment present in a colourless reduced form in areas which have been subjected to ultraviolet radiation. This accounts for the fact that different sources of light—sun, carbon-arc lamp, or mercury-vapour lamp—cause slightly different pigmentation of the skin. The effects of exposure to sunlight are a combination of the effects of heat and ultraviolet light.

It is well known that if unpleasant effects are to be avoided, exposure to sunlight must be gradual. The tanning, which is due to the appearance of pigment in the basal layers of the germinative zone, cannot satisfactorily explain the tolerance to ultraviolet radiation, since in sunburn it is the more superficial prickle-cell layer which is damaged. The stratum corneum passes only a small amount of the ultraviolet radiation which falls on it and its thickening following exposure to sunlight is the main cause of acclimatization and explains the fact that blond persons with little pigment in the skin acquire tolerance to ultraviolet radiation. The dark-skinned races need more ultraviolet radiation to produce an erythema than the white races. This relative immunity to the harmful effects of sunlight is not due to any difference in thickness in the stratum corneum but to the possession by dark races of a greater amount of pigment in this layer which absorbs ultraviolet radiation. The vasodilatation which occurs in sunburn must be an indirect effect since the filter formed by the horny layer probably prevents any ultraviolet radiation reaching the corium which contains the blood vessels. It may be that a substance like Lewis's 'H' substance is elaborated in the epidermis and travels down to the minute vessels in the skin to produce the dilatation.

FAINTING (SYNCOPE)

Syncope is a sudden, transient loss of consciousness due to an inadequate supply of blood to the brain. It may therefore, though rarely, have a *cardiac* cause, as for example when a gross disorder of conduction leads to prolonged diastole of the ventricles.

On the other hand, syncope may have a vasomotor origin. It is then usually called a faint and is accompanied by extreme pallor, nausea, sweating, slowing of the heart, and a fall of arterial pressure to a systolic level of 60 mm Hg or less. A faint rarely lasts more than a few minutes and almost never occurs when a person is horizontal. The fall of arterial pressure is due to dilatation of blood vessels, especially in skeletal muscles, and the cardiac slowing is a vagal effect since it can be abolished by atropine. Because of the association of these two circulatory phenomena the condition was referred to by Lewis as a *vaso-vagal attack*. Syncope is commonly produced by emotional disturbances such as distress, fear, and disgust, and is rarely the result of heart disease. It frequently occurs from the sudden loss of fairly large amounts of blood. In blood donors fainting is uncommon when less than 400 ml. of blood are removed, but occurs in about half the subjects if 1,000 ml. are withdrawn. Loss of blood is at first normally accompanied by an increase in heart rate and vasoconstriction so that the arterial pressure is little altered. As the loss continues, however, syncope occurs quite suddenly with a rapid fall of arterial pressure and cardiac slowing.

A rather rare type of vaso-vagal attack has already been mentioned (p. 512) as being due to stimulation of the carotid sinus. A third form of attack, called *postural syncope*, occurs in persons who stand erect for any length of time without moving, as do soldiers on parade. Gravity and the absence of the 'muscle pump' retard the return of blood to the heart and the cardiac output becomes inadequate to supply the brain. Minor forms of postural syncope amounting to no more than momentary dimness of vision and unsteadiness are fairly commonly experienced when a person rises rapidly from a sitting position, especially when the peripheral vessels are dilated as for example after a hot bath or in an overheated room. A rather similar phenomenon (*effort syncope*) may occur after a sudden strenuous effort especially if it involves the abdominal muscles such, for example, as occurs in a tug-of-war. Blood is driven out of the large veins in the abdomen by the sudden muscular effort, thus leaving a potential reservoir which fills with blood immediately the muscles are relaxed, and the return of venous blood to the heart is temporarily retarded. Syncope may occur during a paroxysm of coughing (*cough syncope*) because the prolonged increase of intrathoracic pressure prevents the return of venous blood to the heart.

PERIPHERAL CIRCULATORY FAILURE (SHOCK)

Peripheral circulatory failure or shock is the term given to the condition that may develop shortly after sudden injury causing severe loss of blood or damage to muscular tissue or after severe burning accidents. The patient is pale and greyish in colour. His skin is cold and shows drops of sweat. The pulse at the wrist may be of very poor amplitude, but is usually rapid at first, although in severe cases it may be slow. Arterial pressure is low. The breathing may be shallow and rapid. The patient is thirsty, he may feel cold, but usually has little pain or interest in his surroundings. All bodily activities are depressed; the metabolic rate is reduced and the temperature low.

Circulatory Alterations.—The vascular system adapts itself to the smaller blood volume by vasoconstriction and the heart rate, cardiac output and circulation rate may be increased. After severe hæmorrhage the systolic blood pressure may fall to 60 mm Hg. The blood volume is greatly reduced at first but it rises slowly by entry into the blood stream of fluid from the tissues, the restoration becoming complete within two to four days after the accident. During this process of *hœmodilution* the concentrations of hæmoglobin and plasma protein fall.

Severe trauma to muscles in man, or in experimental animals, may produce a sudden loss of 30 or 40 per cent. in blood volume, as judged by the dye method (p. 376), which in an average man corresponds to a loss of about 2 litres of blood. Similar rapid reductions of blood volume without actual loss of blood occur consistently after severe burning or following abdominal injuries complicated by peritonitis. This *hæmoconcentration* or *oligæmia* is due to loss of plasma and fluid into the injured area and is marked by an increase in the concentration of hæmoglobin and plasma protein. Here again the capacity of the vascular system is reduced by vasoconstriction to accommodate the smaller amount of blood. The diminished plasma volume is then made good by drawing upon the extravascular fluids in other parts of the body which, however, contain little protein so that the plasma protein concentration falls. When the loss of blood or plasma is too great to be compensated for by vasoconstriction the pressure within the vascular system falls, the cardiac output declines to less than 50 per cent. of the normal and the arteriovenous oxygen difference is much increased. Peripheral circulatory failure (shock) is now said to be present. Low blood pressure by itself is apparently not the cause of shock because low levels of arterial pressure can be maintained in man by drugs (for example hexamethonium bromide) without shock occurring. It is believed to be the result of a considerable and usually sudden, reduction in effective circulating blood volume from hæmorrhage, from excessive permeability of the blood vessels or from dilatation of large areas of the capillary bed. Circulatory failure may also occur as a result of loss of fluid and electrolytes from the body following severe vomiting and diarrhœa.

From what has already been said it follows that in peripheral circulatory failure the circulation is retarded. Stagnation of the blood in peripheral tissues is in fact such a marked feature of shock that the condition has been described as one in which the animal has "bled into its own dilated blood vessels." In these circumstances the inhalation of oxygen clearly cannot augment the oxygen uptake in the tissues; the resulting stagnant anoxia may lower the pH of the arterial blood to 7·0 or even 6·9.

Nervous Factors.—Nervous influences in the production of peripheral circulatory failure have been difficult to demonstrate. If the condition is produced in an animal by simple hæmorrhage from an artery the blood volume is found to be much lower than in an animal in which a similar state has been produced by trauma. It seems that the large number of afferent impulses, produced as a result of the traumatization, lead to the development of peripheral circulatory failure more rapidly than would be expected from the reduction of blood volume alone.

Toxic Factors.—Although 'toxic' factors have been thought to play some part in the production of peripheral circulatory failure it is extremely difficult to furnish experimental proof of their influence. In animals many of the features of shock can be imitated by injecting histamine intravenously but, although it is difficult to get evidence of the accumulation of histamine in the blood after traumatizing muscles sufficiently to produce peripheral failure, an increase of blood histamine after burning has been reported. However, certain species of mammals which are resistant to injection of histamine readily develop all the signs of shock after trauma. The evidence, unsatisfactory though it is, is rather against the idea that histamine is *the* substance responsible for shock. After adrenalectomy animals die in a condition resembling the circulatory failure of shock. Administration of adrenal cortical hormone has, however, not been found useful in treating peripheral circulatory failure in man. Various vasodilator substances, for example adenosine triphosphate, are liberated when muscles are damaged, but it is not easy to decide their importance in producing shock.

Peripheral circulatory failure may occur in the course of severe infections, such as pneumonia. It is not at all clear, however, whether it is due to 'paralysis' of the vasomotor centre, to an effect of the infection on the arteriolar walls, or to dehydration (p. 635).

Despite many studies devoted to this phenomenon in the last two wars, we have still no satisfactory explanation of peripheral circulatory failure, or any infallible method of treating it. Animal experiments have clearly shown, however, that if it is allowed to persist beyond a certain time it becomes irreversible.

REFERENCES

Barcroft, H. & Swan, H. J. C. (1953). Sympathetic control of human blood vessels. *Monographs of Physiological Society No. 1.* London: Arnold.

Clark, E. R. (1938). Arterio-venous anastomoses. *Physiol. Rev.* **18,** 229–247.

Dawes, G. S. & Comroe, J. H. (Jr.) (1954). Chemoreflexes from the heart and lungs. *Physiol. Rev.* **34,** 167–201.

Giese, A. C. & Wells, J. M. (1946). Sunburn protection, natural and artificial. *Sci. Mon.* **62,** 458–464.

Grant, R. T. & Reeve, E. B. (1951). Observations on the general effects of injury in man, with special reference to wound shock. *Spec. Rep. Ser. med. Res. Coun., Lond., No. 277.* Pp. 1–313. London: H.M.S.O.

Harrison, T. R. (1938). Arterial and venous pressure factors in circulatory failure. *Physiol. Rev.* **18,** 86–108.

Lewis, T. (1927). *The Blood Vessels of the Human Skin and their Responses.* London: Shaw & Sons.

Miner, R. N. (Ed.) (1952). The shock syndrome. *Ann. N.Y. Acad. Sci.* **55,** 345–542.

Peripheral circulation in man (1954). *Ciba Foundation Symposium.* London: Churchill.

Sharpey-Schafer, E. P. (1952). Peripheral circulation in circulatory failure. *Brit. med. Bull.* **8,** 331–332.

Sjöstrand, T. (1953). Volume and distribution of blood and their significance in regulating circulation. *Physiol. Rev.* **33,** 202–228.

Uvnas, B. (1954). Sympathetic vasodilator outflow. *Physiol. Rev.* **34,** 608–618.

Visceral circulation (1952). *Ciba Foundation Symposium.* London: Churchill.

CHAPTER 31

CIRCULATION THROUGH LUNGS, LIVER, SPLEEN AND BRAIN

PULMONARY CIRCULATION

THE lungs, like the liver, have two blood supplies which anastomose to some extent. The pulmonary artery supplies the respiratory bronchioles and their subdivisions down to the alveoli; this area of the lung which is concerned with gaseous exchange is drained by the pulmonary veins. The bronchial arteries, branches of the aorta, supply the bronchi and all their subdivisions down to the respiratory bronchioles. There is, then, an anastomosis between the systemic and pulmonary circulations at the level of the respiratory bronchioles. The blood from the larger bronchi drains away by the bronchial veins but the blood from the smaller bronchi and the respiratory bronchioles drains into the pulmonary veins. In this way the large pulmonary veins contain chiefly oxygenated blood from the pulmonary circulation but a certain amount of venous blood from the bronchial circulation.

In addition to anastomoses between the pulmonary and bronchial vascular trees which are probably extremely complicated, the lungs, like other tissues, contain numerous arterio-venous shunts (10 to 500 μ in calibre) between the pulmonary arteries and veins which no doubt play an important role in the physiology of the pulmonary circulation, particularly when part of a lung is the seat of disease. For example, when a lobe of a lung is completely airless as the result of pneumonia, the blood in the systemic arteries may remain fully saturated, suggesting that large amounts of blood in the pulmonary circuit have been short-circuited from the diseased area.

The pressure in the pulmonary artery in man has been measured directly by catheterization through the right side of the heart (Fig. 31, 1). At rest the systolic pressure is only 19 to 26 mm Hg which is of course the same as the systolic pressure in the right ventricle. The diastolic pulmonary artery pressure is from 6 to 12 mm Hg. As the mean pulmonary pressure is about 15 mm Hg the total resistance in the pulmonary circuit (see formula, p. 499) is about one-sixth of that in the systemic circuit and thus the right ventricle has less work to do than the left. When the catheter is advanced to the point at which it will pass no further, oxygenated blood can be withdrawn from it. Its tip must block a small branch of the pulmonary artery and the pressure recorded, 6 to 12 mm Hg, is therefore taken as the pressure in the pulmonary capillaries. It is much below the pressure in the systemic capillaries although of course the osmotic pressure exerted by the plasma protein is the same in the two territories. In other words the tendency for fluid to remain within the pulmonary capillaries or for fluid to enter them from the alveolar spaces is greater than in the systemic capillary bed. When the pulmonary capillary pressure rises to

20 to 30 mm Hg, as it does in cardiac failure, this property is lost. The low resistance to flow in the pulmonary circuit accounts for the fact that the mean pressure difference between the pulmonary artery and the left atrium is only about one-fifth of that between the aorta and the right atrium. The low resistance is accounted for by peculiarities of structure of the blood vessels of the lungs; the arterioles contain relatively little smooth muscle and the capillaries and veins are very large, short, and readily distensible. The low peripheral resistance and the large blood flow mean that the time taken for blood to pass through the pulmonary vascular bed (pulmonary circulation time) is very short; in fact it has been found to be about 5 seconds.

These facts allow the amount of blood in the pulmonary circuit to be simply calculated. If 4·8 litres are pumped through the lungs per minute then in 5 seconds $5/60 \times 4{\cdot}8 = 0{\cdot}4$ litre pass completely through the circuit. Under basal conditions the pulmonary circuit must therefore contain some 400 ml. or approximately 10 per cent. of the total blood volume. The properties of the pulmonary vascular bed are such that in healthy persons this can be increased to accommodate about 20 per cent. of the total blood volume with comparatively little increase of pressure. When some 420 ml. of blood is withdrawn from healthy persons the sizes and numbers of blood vessels in the peripheral parts of the lungs is reduced, as determined radiographically. The pulmonary vascular bed may therefore be regarded as a reservoir of blood in the sense that it can accommodate more or less blood as required. Anoxia, however, produces vasoconstriction in the pulmonary circuit, and structural changes in the lungs that develop as the result of disease lead to diminished distensibility. Both factors favour an increased pulmonary vascular resistance which, in experiments on animals made temporarily anoxic, can be so great as to lead to a 60 per cent. reduction in flow. Autonomic nerve impulses may be responsible for the increased resistance since it may be reduced in man by autonomic blocking agents.

It has been estimated that blood remains in the alveolar capillaries for about 0·75 seconds at rest and for approximately 0·3 seconds during strenuous exercise; during this short period gas exchange must take place.

Fig. 31, 2 shows the variations in pressure in the pulmonary system that occur with respiration. With the fall of intrathoracic pressure during inspiration there is increased filling of the right ventricle and therefore a larger systolic discharge. The pulse pressure in the right ventricle is increased and the mean pressure in the pulmonary artery rises. The expansion of the lungs leads to an increased resistance to flow through the capillaries and a smaller quantity of blood leaves by the pulmonary veins than is received through the pulmonary artery. The inflow to the left side of the heart is reduced and therefore the output from the left ventricle is temporarily reduced. The pulse pressure in the femoral artery therefore falls and the mean arterial pressure is reduced. In expiration, relaxation of the lungs improves the venous return along the pulmonary veins, the output of the left ventricle is increased and the pulse pressure and mean arterial pressure in the femoral artery are elevated. Simultaneously the higher intrathoracic pressure reduces the return to the right ventricle and there is a fall in pulse pressure and mean pressure in the pulmonary circuit. This point was referred to on p. 497.

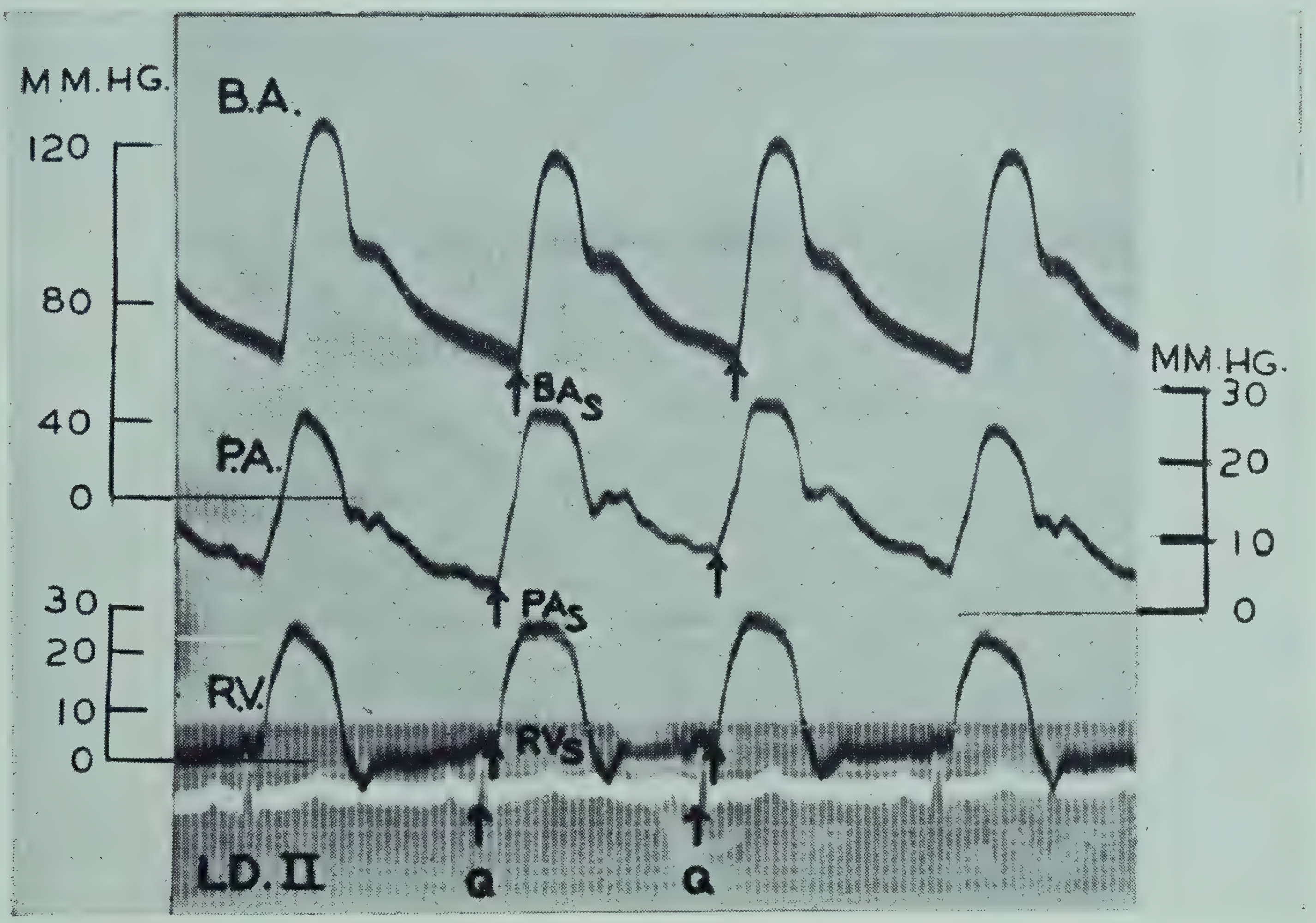


FIG. 31, 1

Simultaneous records of pressure in brachial artery, B. A., pulmonary artery, P.A., and right ventricle, R.V., obtained by optical manometers attached to cardiac catheters. An electrocardiographic record, lead II, is also shown (LD. II).

(*By courtesy of A. Cournand.*)

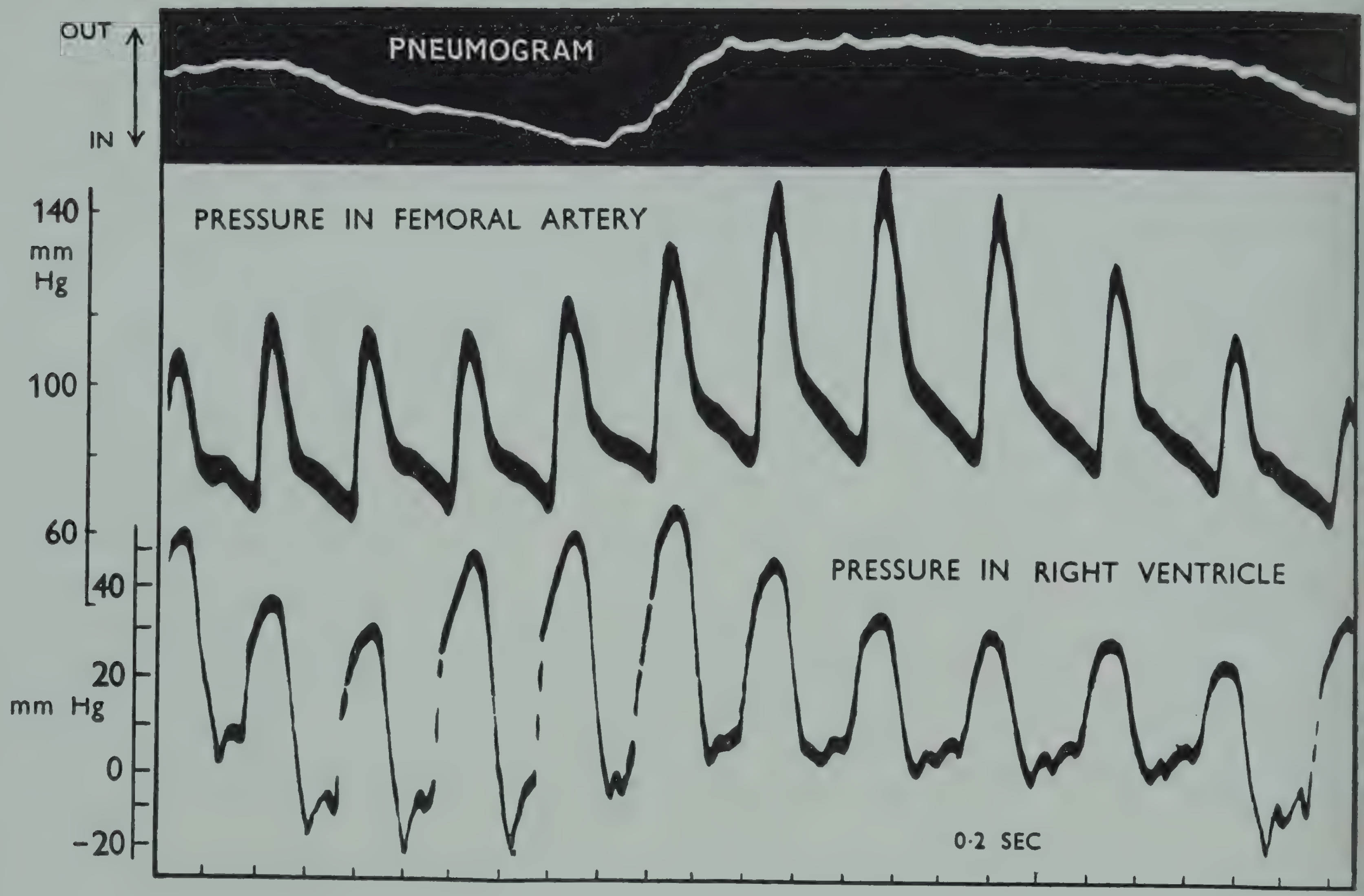


Fig. 31, 2

Simultaneous records of respiration (top trace), blood pressure in femoral artery (middle trace) and pressure in right ventricle (lowest trace) during deep breathing. (*After* A. Cournand (1947), *Bull. N.Y. Acad. Med.* **23**, 34.)

In exercise there is considerable dilatation of the pulmonary vascular bed. This so reduces the resistance in the pulmonary circuit that the mean pressure in the pulmonary artery in man increases only slightly during exercise in spite of a large increase in cardiac output.

The blood vessels of the lungs have a sympathetic supply and a parasympathetic supply from the vagus but the results of stimulation in animal experiments are variable. Adrenaline and noradrenaline usually produce vasoconstriction and acetylcholine vasodilatation. The role of the nerves in the intact animal is not known.

CIRCULATION THROUGH THE LIVER

Attempts to measure the blood flow through the liver have been made in man by applying the Fick principle (p. 449). Bromsulphalein (p. 421) is infused intravenously at a rate such that its concentration in the blood remains constant. Blood is then withdrawn from the hepatic vein through a catheter introduced into it from an antecubital vein. By this method it has been found that the hepatic blood flow in man is about 1·5 litres per minute (100 ml. per 100 g. per minute), a figure which corresponds to results obtained with direct measurements in the dog and which is about half the flow in the inferior vena cava. About three-quarters of this large flow reaches the liver through the portal vein and the remainder through the hepatic artery; in other words the blood supplying the liver is largely venous blood. That the relatively small contribution of arterial blood from the hepatic artery is, however, of vital importance is indicated by the fact that ligation of this vessel in many animals and in man is fatal, the inevitable result being massive destruction (necrosis) of the liver due to the multiplication in it of anærobic bacteria.

The blood of the portal vein drains into the interlobular veins at the periphery of the liver lobules and then passes, by way of numerous sinusoids, between the columns of liver cells to the central (intralobular) veins in the centre of the liver lobules, whence it flows along the hepatic veins to the inferior vena cava (see Fig. 31, 3). The hydrostatic pressure in the hepatic veins is about 6 mm Hg and the blood only about 67 per cent. saturated with oxygen. The blood of the hepatic artery, after supplying the connective tissue of the hepatobiliary (Glisson's) capsule, drains either into venules which join the portal system or into the sinusoids near the margin of the liver lobules. Direct connexions exist, however, between branches of the hepatic artery and the intralobular veins. The presence of the anastomosis between the hepatic and portal systems is shown by the following experiments. If the mesenteric artery supplying the intestine is tied there is only a slight fall of pressure in the portal vein; ligature of the hepatic artery alone produces a fall of portal pressure.

The portal venous pressure has been measured in man by putting a needle connected to a manometer directly into the portal vein at operation. It can also be measured in the conscious subject by inserting a fine needle into the splenic pulp from outside for the intrasplenic pressure is only slightly higher than the portal venous pressure (p. 531). Alternatively, a fine catheter (cardiac catheter) can be passed from the brachial vein through the right atrium until its tip blocks a small radicle of the hepatic vein within the liver. The pressure then recorded approximates to the intrasinusoidal pressure and hence the pressure in the portal vein. By these means the portal venous pressure in man has been found normally to be some 5 to 15 mm Hg. When there is obstruction to the outflow of blood from the portal vein (*portal hypertension*), this may increase to 30 or 40 mm Hg. Since the hepatic blood flow is large and the portal blood pressure low, the resistance to flow through the liver must be very small.

The portal vein is formed by the union of the superior mesenteric vein, bringing blood from a large part of the alimentary tract, and the splenic veins conveying blood (about 45 per cent. of the total) from the spleen, pancreas, gall-bladder and stomach. It has been shown by injection studies in several species, however, that the blood from the superior mesenteric vein does not mix in the portal vein at all freely with blood flowing from the splenic vein. The result is that blood from the superior mesenteric veins enters mainly the right lobe of the liver, while the splenic blood is distributed mainly to the left lobe of the liver. This curious feature of the blood flow in the portal vein may account for the fact that disease, especially in experimental animals, may be confined to one or other lobe of the liver.

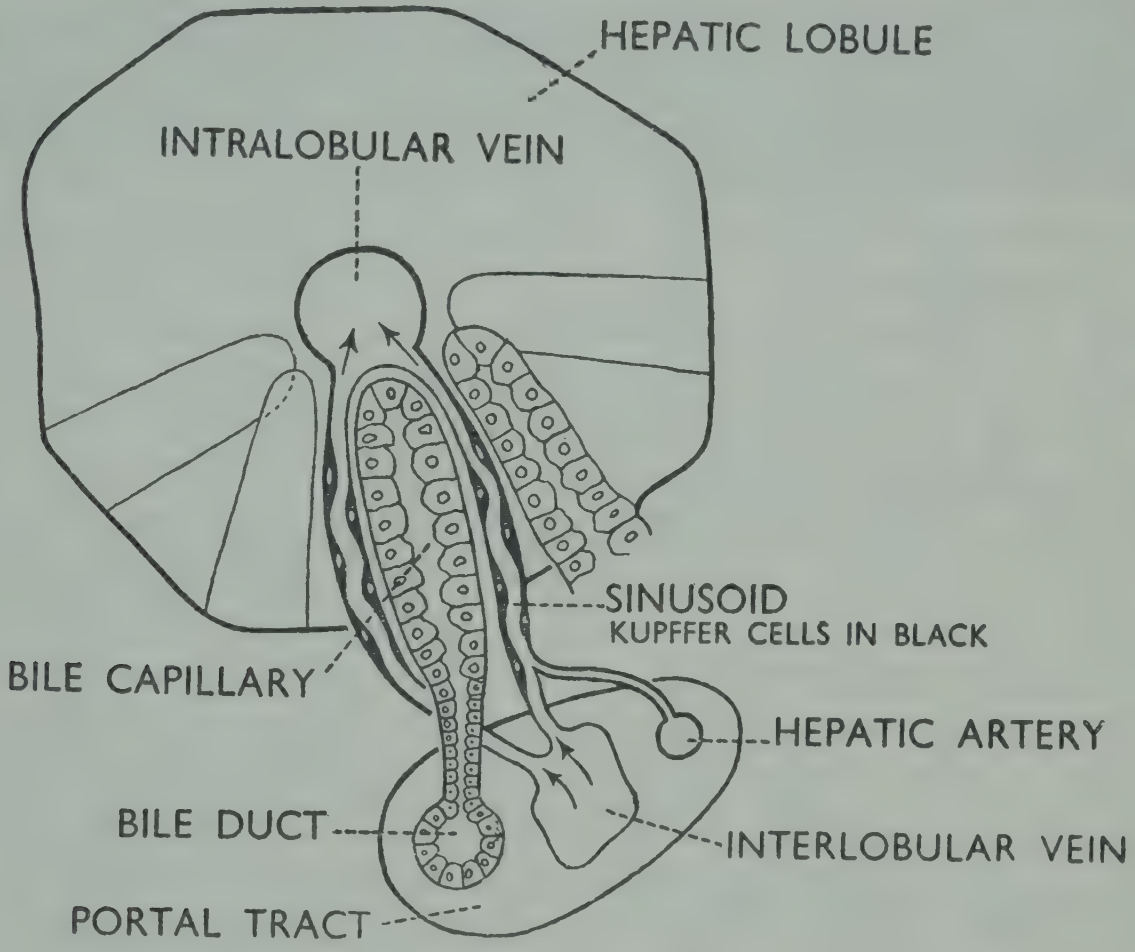


Fig. 31, 3

Structure of liver lobule. Arrows indicate direction of blood flow. (*After* J. W. McNee.)

The capillaries and venous tributaries of the portal vein hold an immense amount of blood and constitute a reservoir with a potential capacity of up to one-third of all the blood in the body. When the splanchnic sympathetic nerves are stimulated in an animal there is first a fall of portal pressure due to diminished flow through the constricted mesenteric arteries, followed by a rise to nearly double the basal level, which is due partly to the damming up of blood by the contraction of the small portal venules within the liver and partly to a rise of hepatic artery pressure following the rise in systemic pressure resulting from the splanchnic vasoconstriction.

When the liver is removed from an animal and perfused via the portal vein the flow soon stops even if the perfusing pressure is raised. The obstruction is due to the contraction of a layer of smooth muscle around the outlet of the hepatic veins. This sphincter relaxes after an injection of oxytocin (Chapter 53) or adrenaline, or after stimulation of the sympathetic, but contracts when histamine is injected. It is not known if this sphincter is of importance in regulating the blood flow through the liver in the intact animal; it is doubtful if it exists at all in man.

THE CIRCULATION THROUGH THE SPLEEN

The framework of the spleen consists of trabeculæ of white and yellow elastic tissue together with smooth muscle fibres which, though a prominent feature in the dog, pig and cat, are rather scanty in the human spleen. The spaces between the trabeculæ are filled with the splenic pulp which is a reticulum of delicate fibrous tissue containing large numbers of cells. These are partly blood cells—red cells, lymphocytes and a few granulocytes —and partly cells belonging to the spleen itself, namely splenic or pulp cells, branched reticulum cells and giant cells. The splenic cells, which are phagocytic and belong to the reticulo-endothelial system, have a single large nucleus and often contain red cells in various stages of disintegration. The giant cells which are present in young persons may also contain red cells.

The splenic artery enters at the hilum and breaks up into numerous branches with few anastomoses (end-arteries). Near their terminations the arterioles are surrounded by a sleeve of lymphoid tissue (Malpighian bodies). It is difficult to trace the pathway of the blood through the spleen from this point but probably the arterioles, after a terminal dilatation, link up in a loose way with sinuses. The walls of these sinuses, which are incomplete, are made up of rod-like cells with narrow chinks between them lying parallel with the long axis of the sinus; occasional circularly placed bands of tissue keep the structure together like the hoops of a barrel (Fig. 31, 4). The blood seems to pass from the sinus through the chinks in the wall into the pulp and back again into the sinus but this circulation must be very slow indeed, since in rats slowly poisoned with carbon monoxide the carbon monoxide content of the blood of the splenic pulp remains for a long time lower than that of the blood in the general circulation. The sinuses either link up with one another or join a venule which has a definite complete wall. The veins unite in the hilum to form the splenic vein which is the largest radicle of the portal vein. There is considerable divergence of opinion among experts as to the detailed histology of the spleen and the structure and behaviour of the organ differ considerably from one species to another. In man the intrasplenic pressure obtained by connecting a manometer to a needle inserted through the skin into the spleen varies between 3 and 17 mm Hg. This agrees well with other methods of measuring portal pressure and suggests that the splenic pulp in man is in free communication with the splenic vein.

The spleen can be brought through the abdominal wall of the dog without disturbing its vascular and nervous connexions and made to lie between the skin and superficial muscles so that its movements can be observed in the unanæsthetized animal. It shows rhythmic contractions which occur at a much slower rate than the respiratory rate. Stimulation of the splanchnic nerves, exercise, asphyxia and hæmorrhage all cause a contraction of the organ which adds a considerable number of red cells to the circulation. In the cat and dog a reflex contraction of the spleen occurs when the pressure in the carotid sinus is lowered by occlusion of the carotid artery.

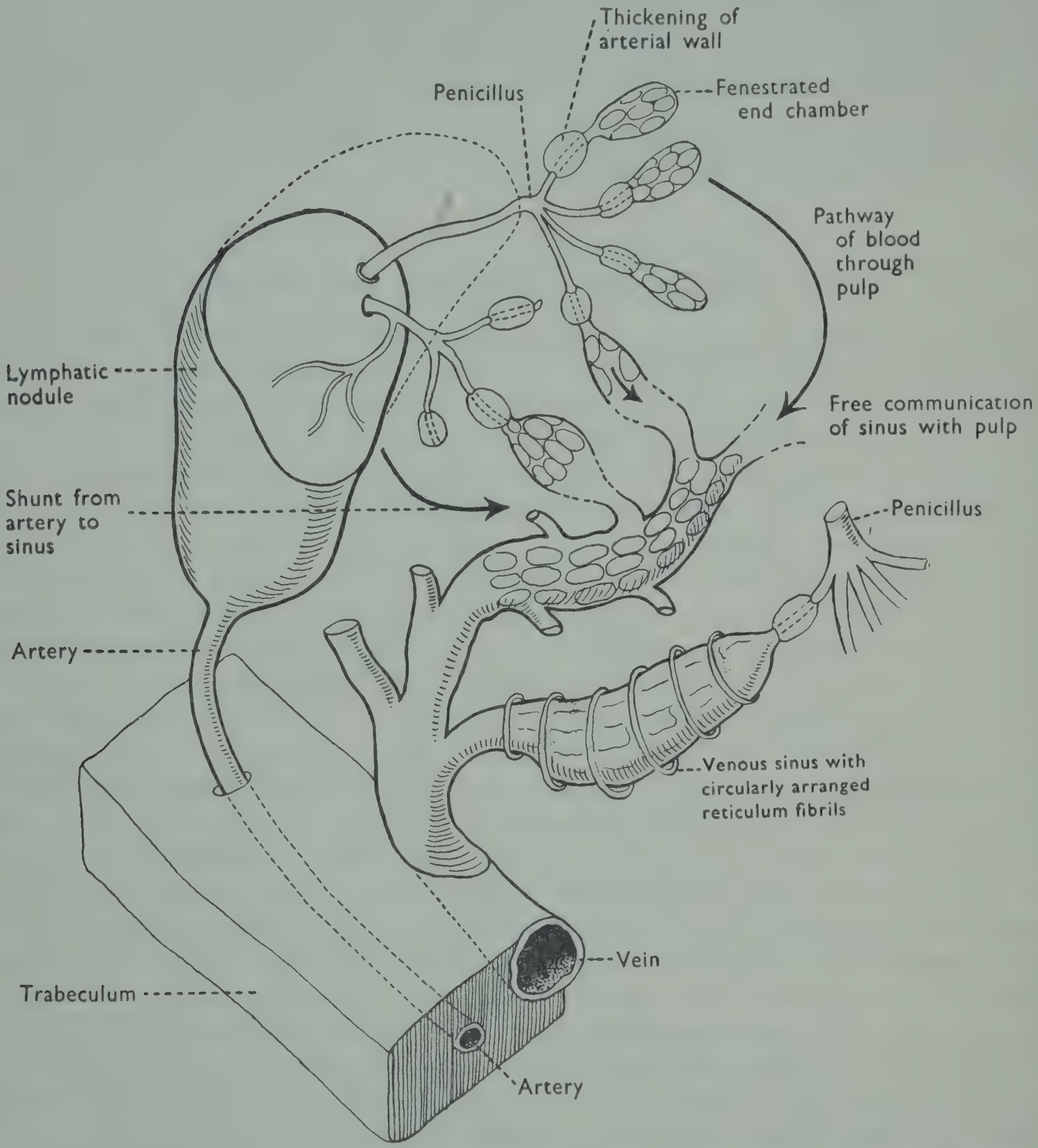

Thickening of arterial wall
Penicillus
Fenestrated end chamber
Pathway of blood through pulp
Lymphatic nodule
Free communication of sinus with pulp
Shunt from artery to sinus
Penicillus
Artery
Venous sinus with circularly arranged reticulum fibrils
Vein
Trabeculum
Artery

FIG. 31, 4

A schematic representation of the blood vessels of the spleen. The arrows indicate the probable routes for passage of blood from splenic artery to vein (E. B. Krumbhaar (1926), Functions of the spleen. *Physiol. Rev.* **6**, 166).

In man the spleen can be made opaque to X-rays by an injection of thorium dioxide (thorotrast). After the injection of 1 mg. of adrenaline the splenic shadow becomes smaller and there is a rise in the red cell count in the general circulation. However, if 1 mg. of adrenaline is given to a splenectomized man there is a similar rise in the red cell count of the peripheral blood. Exercise also produces rises of the same order in normal and splenectomized persons. These alterations are most likely to be due to an alteration in the distribution of plasma and corpuscles (see p. 378). In neither normal nor splenectomized subjects is there any increase in blood volume. Thus, although the spleen undoubtedly acts as a reservoir of red blood cells in the dog, cat and horse, there is no evidence for such a function in man. In any case, an organ of only 200 g. could hardly be an important blood reservoir in man with a blood volume of say 5·5 litres.

CEREBRAL CIRCULATION

The importance of an adequate blood supply to the brain scarcely needs stressing. Actually the arrest of the cerebral circulation for more than about five seconds is followed by unconsciousness, and cerebral ischæmia of longer than twenty seconds by irreparable damage to the cortical neurones. The rate of blood flow through the brain can be measured in man by causing the subject to breathe nitrous oxide, determining the concentration of the gas in samples of blood taken from the internal carotid artery and the jugular bulb, and then applying Fick's principle. Approximately 900 ml. of blood flows through the brain per minute under resting conditions but there may of course be considerable regional variations of flow from time to time within this total figure. In anxiety or apprehension the cardiac output may rise 40 per cent. but there is no corresponding alteration in the total cerebral blood flow. The brain needs some 240 ml. oxygen per 100 g. per hour, which is not very different from the requirement of muscles during active work, and to supply this about 65 ml. of blood pass through the brain per 100 g. per minute. The grey matter is richly supplied with capillaries but the white matter, not so active metabolically, is less vascular.

In the human infant up to the ninth month after birth the skull is not completely ossified—at each beat of the heart the pulsations of the brain are transmitted to the anterior fontanelle. When the infant cries there is a rise of venous pressure and an increase of intracerebral pressure which bulges the anterior fontanelle. Later in life these alterations in volume cannot occur because the brain is then enclosed within the cranium, but if an opening is made in the skull the brain tends to expand through it. The intracranial pressure can be measured in animals by screwing a watertight window with a side tube into the skull and finding the pressure required to push the brain back to its original position. In the adult, just as in the child, the intracranial pressure rises when violent expiratory efforts are made. It falls in the standing position because venous drainage is facilitated by gravity. Blood entering the skull at each systole displaces a similar amount of blood from the veins, but if there is a prolonged increase of arterial pressure cerebrospinal fluid may also be forced out of the skull.

In normal man, the blocking of all sympathetic connexions with the cerebral vessels by injection of local anæsthetic into both stellate ganglia does not lead to any increase in cerebral blood flow. The cerebral arteries do not participate to any extent in the vascular adjustments initiated through the carotid sinus (p. 464) but the pressure in them is controlled by the carotid sinus mechanism which may be regarded as a device to safeguard cerebral blood flow. Sudden falls of cerebral blood pressure are followed by vertigo (dizziness) and confusion and, if more profound, by loss of consciousness. These phenomena are familiar as a faint (vasovagal syndrome, p. 523). They may appear as the result of a sudden change of position; a particular example of this is standing up suddenly from a hot bath. In this case the skin vessels are widely dilated and a large proportion of the blood is circulating in them. If the peripheral vessels do not contract quickly enough on standing the cerebral arterial pressure falls rapidly.

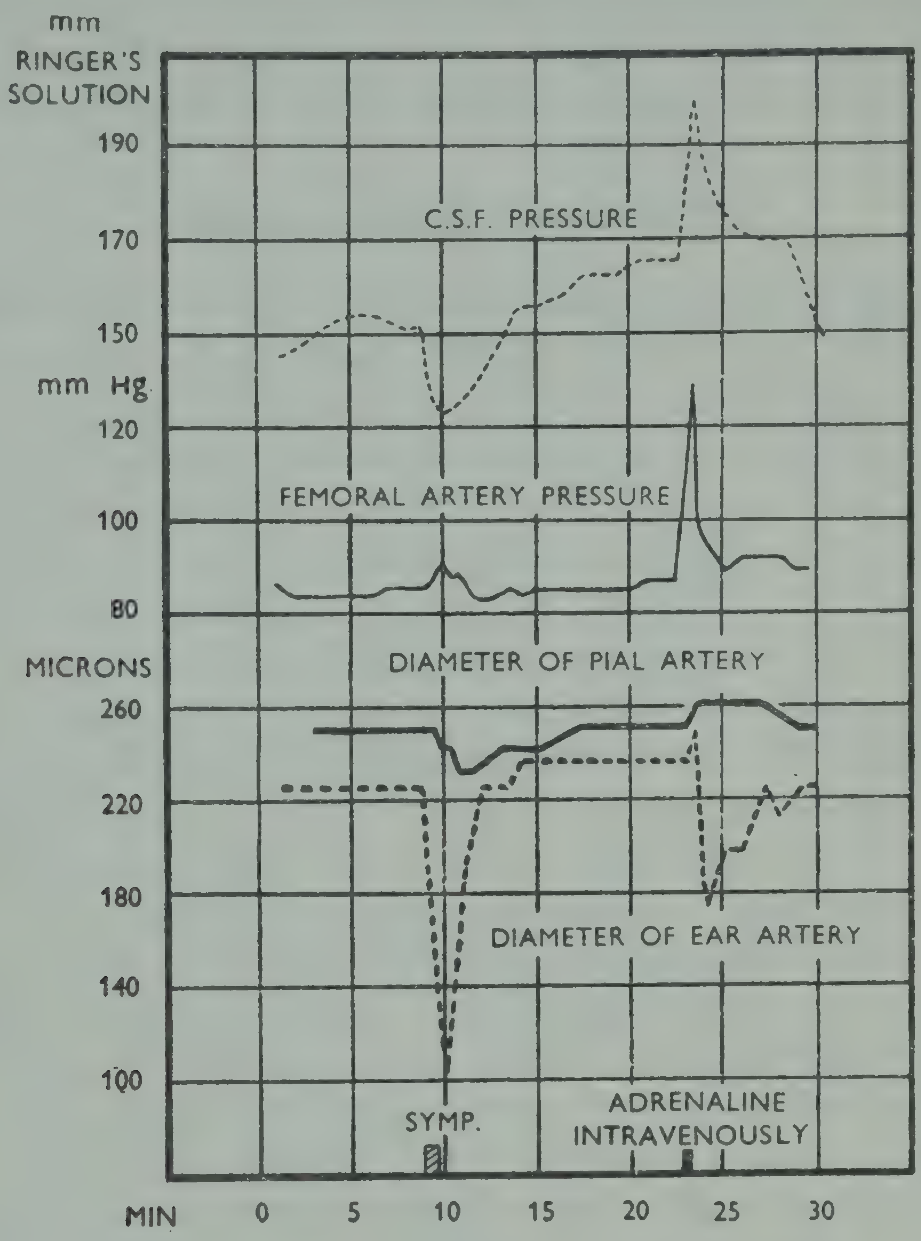


FIG. 31, 5

Sympathetic stimulation and intravenous injection of adrenaline (monkey under dial anæsthesia). The cephalic end of the cut sympathetic nerve (on the left) was stimulated. The pial artery (left parietal region) constricted 7 per cent., as contrasted with a 56 per cent. constriction of the artery in the ear. Twelve minutes later 1 ml. of a 1: 100,000 dilution of adrenaline solution was injected into the femoral vein. The pial artery dilated 4 per cent., while the artery in the ear first dilated 5 per cent. and then promptly constricted 26 per cent. (H. S. Forbes, K. H. Finley & G. I. Nason (1933), Cerebral circulation. *Arch. Neurol. Psychiat. Lond.* **30,** 974).

The pial vessels can be observed by inserting a glass window in the skull and replacing the air below it by Locke's solution. When the skull is maintained as a rigid box, and the pressure relationships are normal, the brain shows no pulsations at all. This is in marked contrast to the obvious pulsations occurring when the skull is incomplete, as in the child, or after trephining. Experiments on animals with cranial windows show that stimulation of the cervical sympathetic produces a slight narrowing of the pial vessels which is, however, very much less than that seen in an artery of similar size in the ear (Fig. 31, 5). This nervous control has not much physiological importance, since an injection of adrenaline is nearly always followed by passive dilatation of the pial arteries owing to the increase of systemic arterial pressure. The pial arteries have a relatively small amount of muscular tissue in their walls compared with a systemic artery of comparable size and it is likely that their diameter depends largely on the internal pressure. Reflex vasodilator effects can be demonstrated by stimulation of the vagus, the efferent fibres passing out in the VII nerve. In asphyxia there is a dilatation of the cerebral vessels which may be due either to the direct action of carbon dioxide or to the rise of systemic blood pressure.

The venous sinuses of the dura carry away the blood from the brain; in addition they also receive cerebrospinal fluid through the arachnoid granulations (p. 646).

REFERENCES

Altschule, M. D. (1949). *Physiology in Diseases of the Heart and Lungs.* Oxford University Press.

Bradley, S. E., Inglefinger, F. J., Bradley, G. P. & Curry, J. J. (1945). Estimation of hepatic blood flow in man. *J. clin. Invest.* **24,** 890–897.

Child, C. G. (1954). *The Hepatic Circulation and Portal Hypertension.* London: Saunders.

Cournand, A. (1947). Recent observations on the dynamics of the pulmonary circulation. *Bull. New York Acad. Med.* **23,** 27–50.

Daly, I. de Burgh (1935–36). The physiology of the bronchial vascular system. *Harvey Lect.* **31,** 235–255.

Schmidt, C. F. (1955). *The Cerebral Circulation in Health and Disease.* Oxford: Blackwell.

Sherlock, S. (1955). *Diseases of the Liver and Biliary System.* Oxford: Blackwell.

Snook, T. (1950). Comparative study of the vascular arrangements in mammalian spleens. *Amer. J. Anat.* **87,** 31–61.

CHAPTER 32

RESPIRATION

IN unicellular organisms oxygen can diffuse from the watery environment to all parts of the cell. With increasing size and complexity of the animal, however, the distance between the surface and the tissues becomes so great that special means are required to carry oxygen to the individual cells. In vertebrates this is achieved by special organs, gills in fishes and lungs in terrestrial animals, which expose a very large surface richly supplied with blood vessels to the oxygen dissolved in the water or to the oxygen within the lungs. The oxygen taken up in the capillaries of the lungs is conveyed by the blood to the tissues. At the same time the carbon dioxide produced in the tissues is carried to the lungs and there leaves the body. A scheme of the processes involved in respiration is given in Fig. 32, 1.

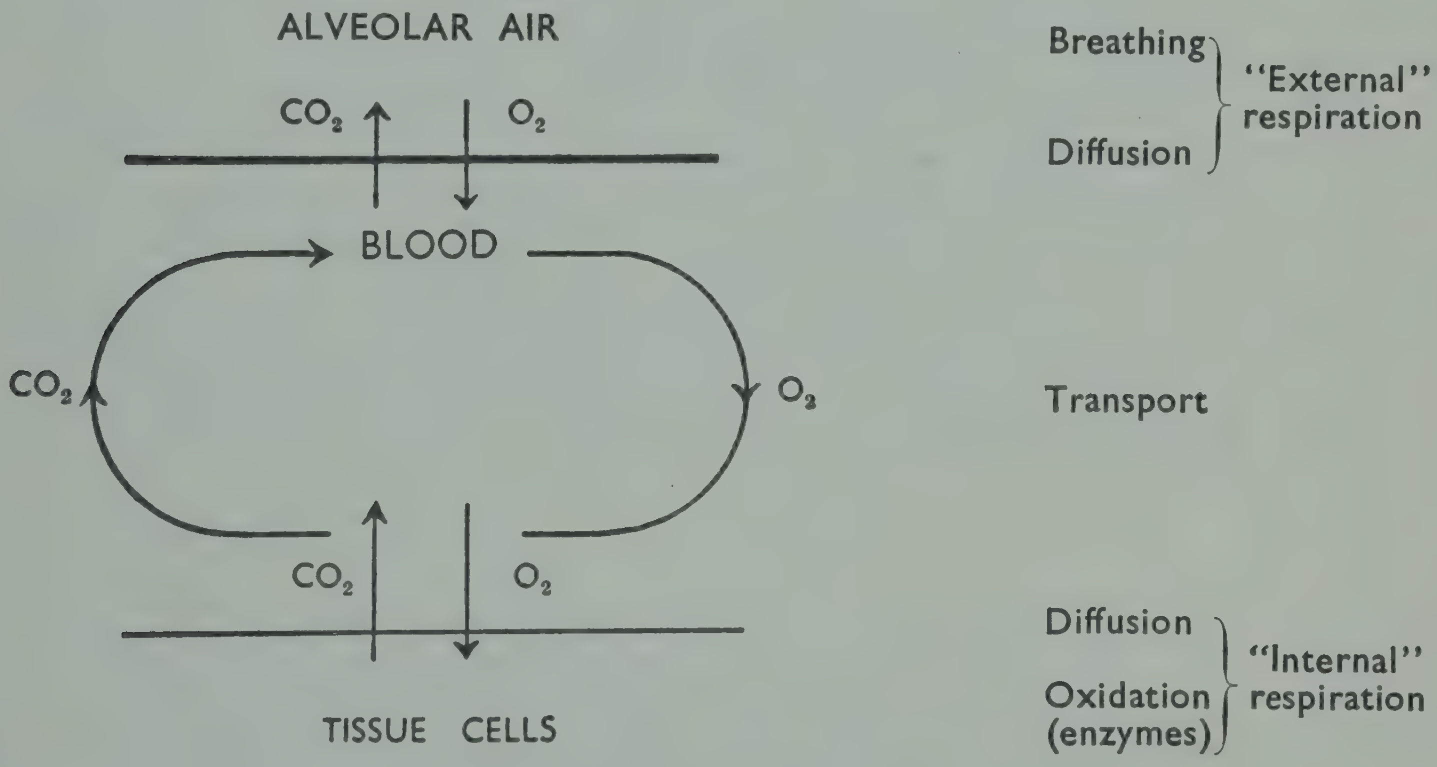


FIG. 32, 1

During breathing, air passes through the nose, and often the mouth, into the pharynx, through the opening between the vocal cords into the trachea, and then into the bronchi and smaller air passages till it reaches the alveoli of the lungs where the exchange of gases occurs.

The Nose.—The nasal passages are lined by a very vascular mucous membrane, characterized, except at the entrance, by ciliated columnar epithelium which warms and moistens the entering air and removes a certain amount of dust. The warming is quite effective as air inspired at 6° C, for example, has been heated to 30° C by the time it reaches the back of the nose. Its temperature rises as it passes down the lungs to body temperature but the temperature of the expired air is always below 37° C as the air loses heat as it passes out through the nose.

When the mucous membrane is congested or inflamed there is considerable resistance to inflow and breathing through the nose is correspondingly difficult.

Exposure of the skin of the face to infra-red rays, from an electric radiator for example, causes a swelling of the mucosa. Short-wave radiation or cooling of the skin produces a shrinkage. Discomfort in artificially heated rooms (stuffiness) may be at least partly explained by the absence of short-wave radiation, and heating systems should ideally be designed to avoid swelling of the nasal mucosa and the discomfort so produced. In monkeys the nasal mucosa becomes hyperæmic at the premenstrual stage of the menstrual cycle and injections of œstradiol into castrate monkeys, both male and female, also produce hyperæmia of the conchæ.

STRUCTURE OF THE LUNGS

The Air Passages.—The *trachea* is a tube 1·5 to 2 cm in diameter and 10 cm long. It is kept open by incomplete rings of cartilage, the deficiency on the posterior surface being filled in with fibrous and muscular tissue. This structure allows free movement of the head and neck without the danger of kinking and obstructing the air way. The trachea divides into

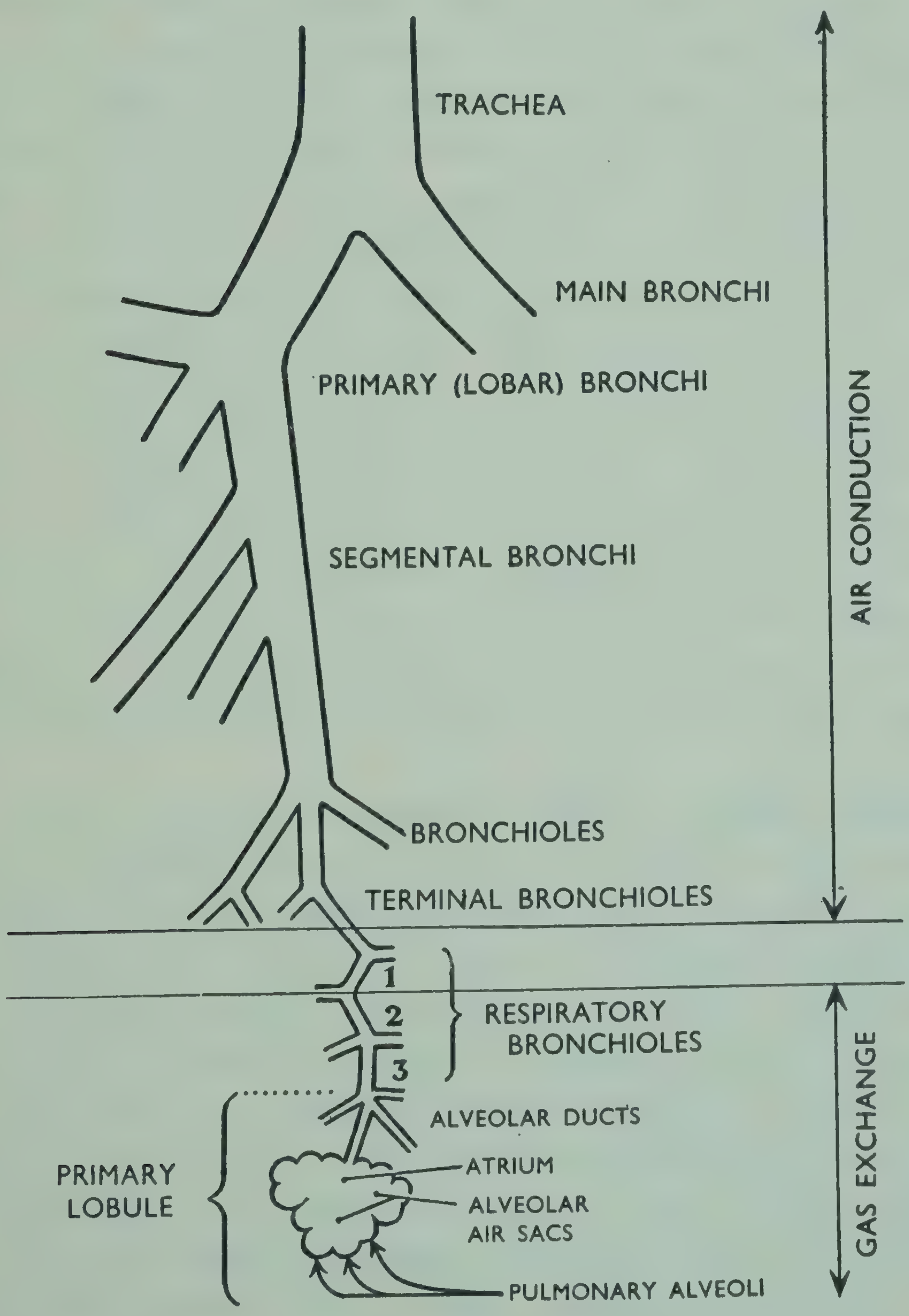

TRACHEA
MAIN BRONCHI
PRIMARY (LOBAR) BRONCHI
SEGMENTAL BRONCHI
BRONCHIOLES
TERMINAL BRONCHIOLES
AIR CONDUCTION
1
2
3
RESPIRATORY BRONCHIOLES
ALVEOLAR DUCTS
PRIMARY LOBULE
ATRIUM
ALVEOLAR AIR SACS
PULMONARY ALVEOLI
GAS EXCHANGE

Fig. 32, 2

The air passages. (*After* Macklin.)

two main *bronchi* and these in turn subdivide repeatedly, the smallest air passages being called *bronchioles.* All these air passages have a fibrous outer layer; in addition the larger bronchi are supported by irregular pieces of cartilage. Internal to this is a layer of smooth muscle, the *bronchial musculature*, and between the muscle fibres there is a considerable amount of elastic tissue. The tubes are lined by a mucous membrane with a ciliated columnar epithelium. In addition in the trachea and larger bronchi numerous mucous glands are found below the epithelium. The terminal bronchioles, the muscle of which is particularly well developed, open into *respiratory bronchioles* of equal diameter (Fig. 32, 3). A small number of these open directly into *alveoli* but for the most part alveolar ducts form the final portion of the respiratory tree, connecting the respiratory bronchioles with the pulmonary alveoli. These number approximately 700×10^6 in man and are about 190 μ in diameter.

Ciliary Activity.—The respiratory tract from the nose to the bronchioles is lined by ciliated columnar epithelium. The free border of these cells shows numerous hair-like processes, or cilia, which appear under the microscope to impel the fluid on the surface in one direction by a rapid forward movement (the effective beat) followed by a slower return movement. This movement can be seen on the surface of a nasal polypus recently removed from the human nose or more conveniently on the surface of the tongue or pharynx of a frog. In the frog, the cilia beat about 10 to 17 times per second and the effective beat is from 5 to 6 times as fast as the return beat so that particles placed on the pharynx are carried slowly along. Particles of powder are carried along the dog's trachea *in situ* at a rate of 0·4 mm per sec. In man the ciliary movement propels mucus and foreign bodies to the main bronchi and trachea, from which they are finally cleared out by coughing. Ciliary activity continues independently of all nervous connexions but it can be arrested by a small dose of X-radiation. Low concentrations of acetylcholine applied to a cilated epithelium increase the rate of passage of particles of powder over it whereas adrenaline may have a depressant action. Whether these substances play any part in the physiological control of ciliary activity is not known.

The Respiratory Parts of the Lungs.—The structures distal to each alveolar duct are spoken of as a *primary lobule* which is the physiological unit of the lung (Fig. 32, 2). The respiratory bronchioles are lined by cubical, non-ciliated epithelium, which becomes progressively flatter as it is traced towards the pulmonary alveoli. The movement of air is aided by minute vents which connect adjacent alveoli within each primary lobule. The existence of these vents in man has been proved by passing a catheter with a wide terminal flange into a bronchus until the bronchus is blocked. The other end of the catheter is connected to a volume recorder. If the catheter is closed during inspiration and opened during expiration the volume recorder rises; this shows that air must have entered the blocked bronchi from the distal end. Lambert has recently demonstrated epithelium-lined tubules connecting terminal or respiratory bronchioles with adjacent alveoli; a communication is thus provided between the ultimate alveoli and their parent bronchioles.

The exchange of gases between the air and blood takes place in the alveoli where, according to classical opinion, the air and the blood are

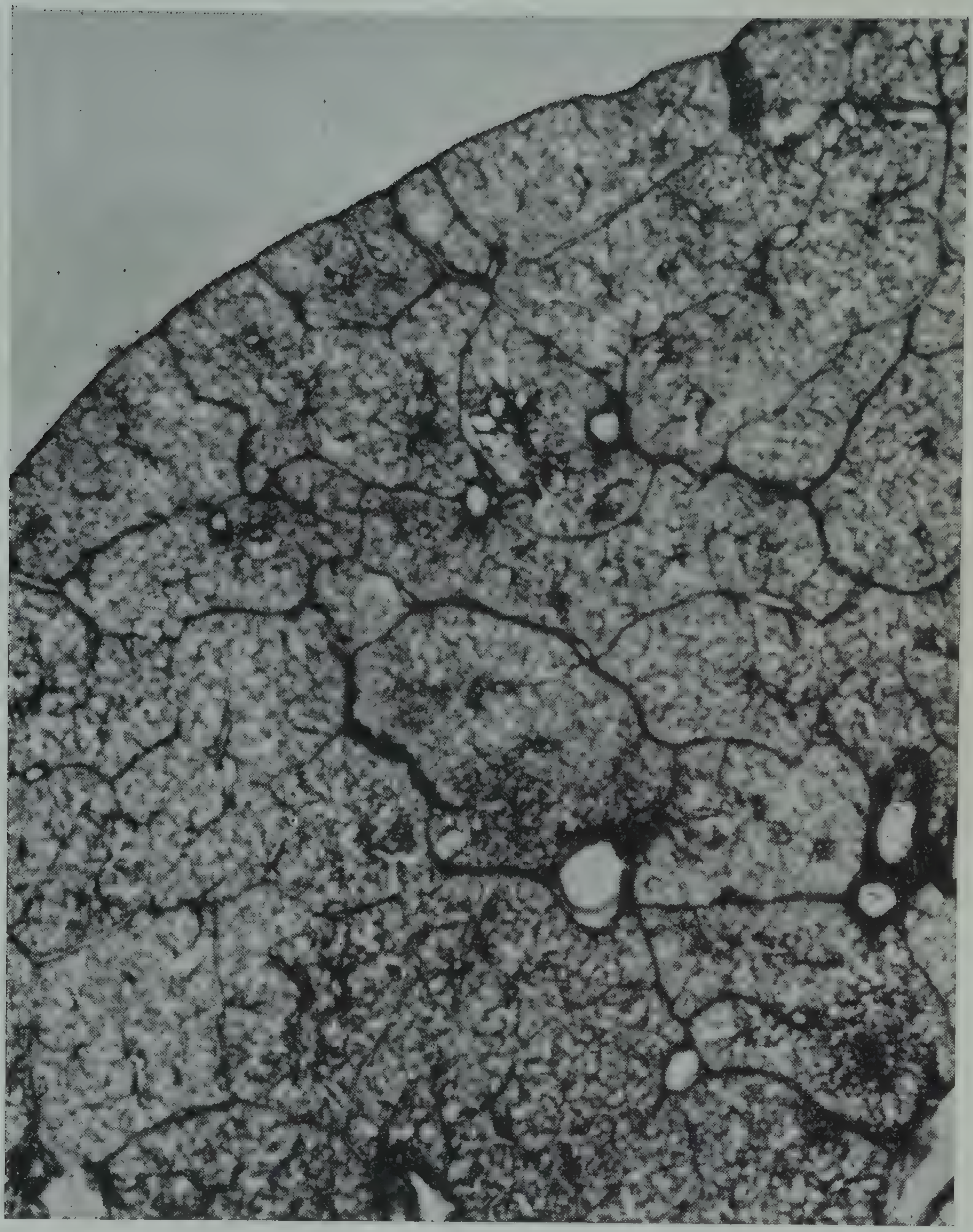


FIG. 32, 3

Photograph (actual size) of part of a section of the lung of a man, aged 60, in which the secondary lobules are outlined by deposits of coal dust in the lymphatics of the interlobular septa.
(*By courtesy of J. Gough.*)

separated by two very thin cellular membranes, the endothelium of the pulmonary capillary and the flattened epithelium of the alveolus. Until the fourth month of intrauterine life the terminal 'air' spaces in the lungs are lined by a layer of epithelium, but about the fifth month of intrauterine life marked changes occur in the histological appearances. The pulmonary capillaries seem to grow into the future air spaces and the alveolar epithelium desquamates, so that at full term little trace of epithelium can be found. Thus according to some investigators only one cellular layer intervenes between blood in the capillaries and air in the alveoli. In any case the thickness of the tissue intervening between the alveolar air and the blood is of the order of 1 μ. The total area of the alveolar membrane of both lungs is estimated to be 50 to 80 sq. m in man; the area of course depends on the state of expansion of the lungs.

The capillaries in the alveolar walls are derived from the pulmonary artery. The bronchial arteries, branches of the thoracic aorta, supply the larger bronchi, the bronchioles and their muscles, and the pleura. In the region of the respiratory bronchioles there is an anastomosis between the two systems so that some of the blood conveyed in the bronchial arteries returns to the heart by the pulmonary veins.

From an anatomical point of view the fundamental unit of lung structure could be considered to be the *secondary lobule* (Fig. 32, 3) from many hundreds of which the lung whole is made up. The secondary lobules are divided from one another by interlobular septa in which run the small pulmonary veins and the majority of the lymphatics. In the centre of each lobule is a system of bronchi together with ramifications of the bronchial artery and branches of the pulmonary artery. Each secondary lobule is thus a lung in miniature with a bronchial tree and a circulation of its own. Arterial blood enters the lobule in the bronchial arteries which ramify through the lobule in close relation to the bronchi. Venous blood in the branches of the pulmonary artery is likewise distributed. After oxygenation the blood passes into tributaries of the pulmonary vein running in the interlobular septa but some blood leaves the lobule in bronchial veins running alongside the branches of the artery. The main lymphatic drainage is in lymphatics in the interlobular septa though there are some lymphatics along the bronchi.

There are many vascular connexions both within the lobules and near the hilum of the lung between the pulmonary and bronchial circulations. These may increase in numbers and in size when the pressure in the pulmonary veins is elevated over long periods as the result, for example, of the obstruction to the circulation produced by a narrowed mitral valve (mitral stenosis).

The Pleura.—The outer surface of the lung is covered by a delicate serous membrane, the *visceral pleura*, which is reflected from the roots of the main bronchi on to the inside of the thoracic walls and upper surface of the diaphragm. This part is called the *parietal pleura*. The surface of each layer of pleura consists of a single layer of flattened cells, and normally the two layers are separated by a thin film of serous fluid, sufficient only to lubricate the surfaces so that they move over one another easily during breathing. So long as this film is intact the two layers cannot be separated except by considerable force. The potential space between the two layers is known as the pleural cavity. When the two layers of pleura are separated by an accumulation of fluid, or the entry of air, the pleural cavity becomes apparent.

Shape of the Chest.—A cross-section through the middle of the adult chest is an ellipse, the transverse diameter of which is greater than the antero-posterior in the ratio of about 1·3 to 1. In the infant, however, the chest is more nearly circular in section, and in patients with emphysema (p. 546), in whom the chest contains more air than usual, the length of the antero-posterior diameter approaches that of the transverse, so that the chest is circular in cross-section or barrel-shaped (Table 32, 4).

TABLE 32, 4

Actual Measurements in cm at the level of the Xiphisternal Joint, of the Circumference, Transverse and Antero-Posterior Diameters of the Chest of Subjects at various Ages and in an Adult with Emphysema (p. 546) to show the Differences in the Shape of the Chest

Age	Chest Circumference cm	Transverse Diameter cm	Antero-Posterior Diameter cm	Ratio Transverse/A.P.
3 months .	37·5	10·9	9·6	1·14
2 years . .	47·8	15·3	12·4	1·23
Adult (healthy)	89·0	29·6	22·3	1·33
Adult (with emphysema)	87·8	27·3	25·4	1·07

Gee, quoted in *Finlayson's Clinical Manual* (1926). London: G. Bell & Sons.

MOVEMENTS OF RESPIRATION

The movement of air into and out of the lungs is brought about by alterations in the size of the thoracic cavity, the lungs following these variations passively. The most important muscle of respiration is the *diaphragm*, a dome-shaped sheet separating the thoracic and abdominal cavities; it is made up of a central tendon with muscle fibres radiating from it to the circumference of the lower parts of the thoracic wall. A less important part, in quiet respiration, is played by the external and internal *intercostal muscles* which occupy the spaces between the ribs. A large number of muscles attached to the ribs, for example all the serrati, pectorals, latissimus dorsi and sternomastoid, may take part in the respiratory movements but they are brought into action only when respiration is deep or difficult.

On inspiration the muscle fibres of the diaphragm contract and this muscular septum descends and increases the length of the thoracic cavity. At the same time the abdominal contents are pressed down and the abdomen bulges by a relaxation of the abdominal muscles. As the thoracic cage enlarges the lungs must also expand because the visceral pleura cannot part from the parietal pleura on the thoracic wall. Air flows through the nose into the lungs because the air in the atmosphere is at that instant at a higher pressure than the air in the lungs. On expiration the diaphragm relaxes, the abdominal muscles regain their tone and push up the abdominal contents and the diaphragm rises up into the thorax again. Electromyograph studies in man show that even in the vigorous expiratory efforts seen in asphyxia the abdominal muscles come into action only at the end of expiration and even then their activity is quite small compared with that seen during skeletal movements involving these muscles; the variations in intra-gastric, that is intra-abdominal, pressure are correspondingly small. In maximum voluntary breathing efforts, such as occur in the vital capacity test, there is considerable expiratory effort by the abdominal muscles with a large rise of intragastric pressure.

Diaphragmatic Movements.—The liver is so much denser than the lungs that its upper edge can be clearly seen on the fluorescent screen. Since the liver is pressed against the diaphragm by the abdominal muscles, the upper edge of its X-ray shadow, which moves up and down on respiration, can be used to mark the level of the diaphragm. The downward movement is about 12 mm or so on quiet inspiration, but on violent inspiration it may move down 70 to 130 mm. The height of the diaphragm, and thus the capacity of the chest, varies with the position of the body—the diaphragm is highest in the supine position, a little lower in the erect position, and still lower when the subject is sitting, because the abdominal muscles are relaxed in this position. Since the cross-sectional area of the thorax increases from above downwards, any given air intake can be accomplished by a smaller diaphragmatic movement the lower the resting position of the diaphragm. The patient with respiratory disease is therefore often most comfortable when sitting up.

Rib Movements. — During inspiration the ribs are raised by the intercostal and other muscles and the sternum is lifted forward. At the same time the ribs rotate. Since the ribs pass downwards and forward from their articulations with the vertebral column, these movements produce an increase in the antero-posterior and transverse measurements of the thorax. In deep inspiration the ribs are raised so that each takes the position previously occupied at rest by the next higher rib. As the lower ribs are longer than the upper, the measurements in any horizontal plane of the thorax are greatly increased in deep breathing. The deep descent of the diaphragm at the same time produces a great increase in thoracic volume. A further increase may be afforded by a slight extension of the vertebral column, especially in the upper part of the thorax. In deep respiration the increase in chest circumference is between 5 and 11 cm; in quiet respiration it is only 1 cm.

In expiration the inspiratory muscles relax and the movements of the thorax and the lungs just described are reversed. During this largely passive phase the ribs fall and untwist, partly by their own elasticity and partly by the contraction of the muscles attached to them.

The classical description of the action of the intercostal muscles is shown in Fig. 32, 5 and 6. Since, in the case of the external intercostals, Q is further than P from the fulcrum F, contraction of the muscle fibre raises rib B. Rib A is raised by the external intercostal muscle above it. In the case of the internal intercostal muscles (Fig. 32, 6) the conditions are reversed; R is further away than S from the fulcrum so rib A is pulled down when the muscle contracts. There is some doubt of the validity of this description and some authors say that both sets of intercostal muscles elevate the ribs; in conformity with this idea is the recent electromyographic demonstration that both the internal and external muscles contract simultaneously on quiet inspiration and that there is no evidence of antagonism between them.

It is not possible to bring about the diaphragmatic and chest movements of respiration separately. Claims to the contrary are based on the erroneous notion that the extent of the movement of the diaphragm is shown by the movement of the abdominal wall. There is, in fact, a close correspondence between chest and diaphragmatic movements. Calculations of the alterations in volume of the thorax brought about by changes of the external circumference and by changes of diaphragmatic height show that when the volume corresponding to the vital capacity (page 548) is breathed out one quarter of the ventilation is due to chest expansion and three quarters to diaphragmatic movement.

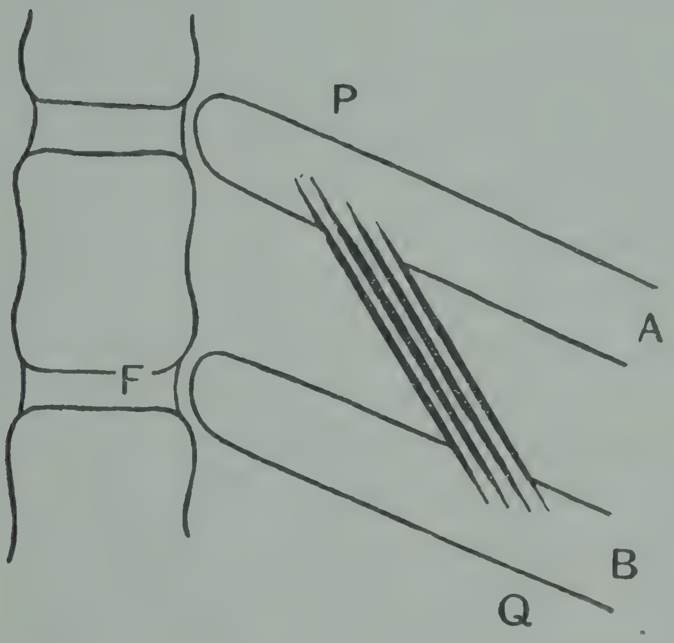


FIG. 32, 5

Diagram to show the action of the external intercostal muscle fibres.

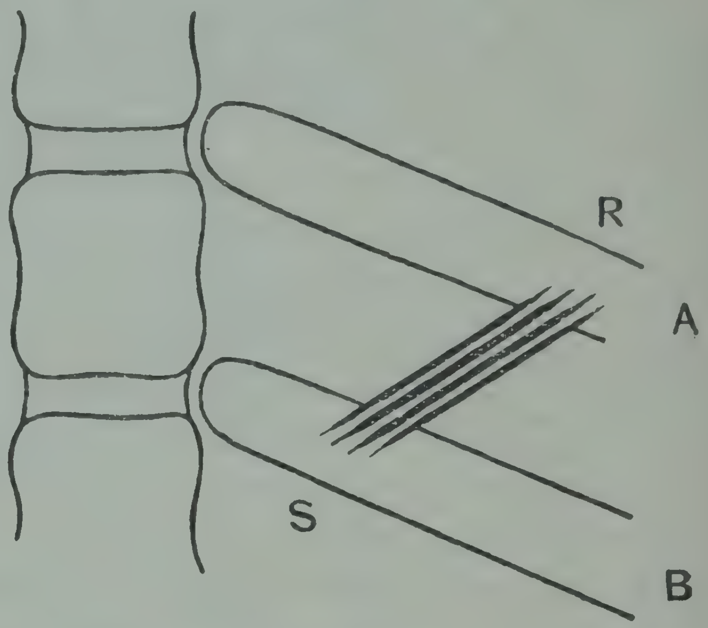


FIG. 32, 6

Diagram to show the action of the internal intercostal fibres.

Expansion of the Lungs.—During inspiration the vocal cords move apart, the bronchi dilate and, as the lung expands, the bronchial tree is elongated, and the alveoli become wider and flatter. At the same time there is, as mentioned on p. 528, an increased amount of blood in the pulmonary circuit. Although alterations in calibre of the bronchi during a respiratory act may be entirely passive it is possible that they depend to some extent on constrictor and dilator impulses passing respectively down the vagus and sympathetic nerves (see p. 546).

Since the lungs expand quite passively, the parts next the most mobile portions of the thoracic cage expand most and the mediastinal portions least. As the lungs expand they occupy the potential space lined by pleura between the diaphragm and the lower ribs (*costo-diaphragmatic recess*), which is obliterated at expiration by the coming together of the two layers of the pleura.

Recording of Respiratory Movements.—The movements of the chest wall are often recorded in experimental work in man and in animals. The simplest way to obtain qualitative records is to use a *stethograph* made of corrugated rubber hose-pipe (respirator tubing) which is tied round the

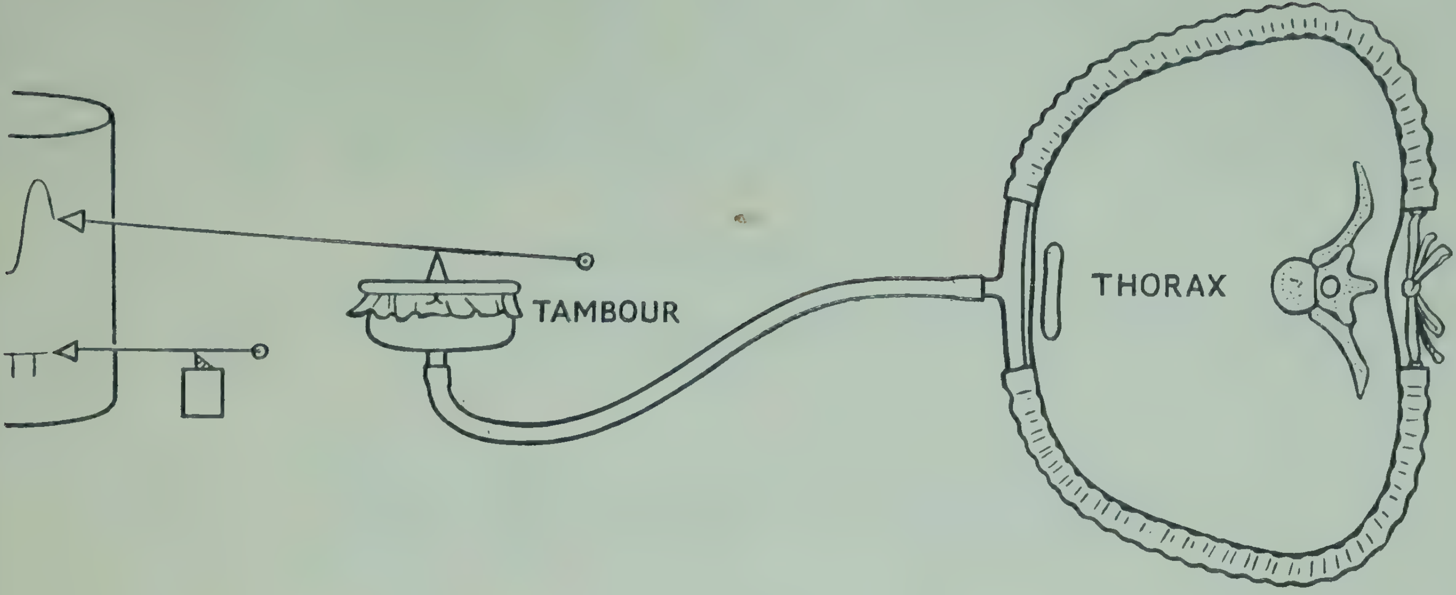


FIG. 32, 7

A simple form of stethograph.

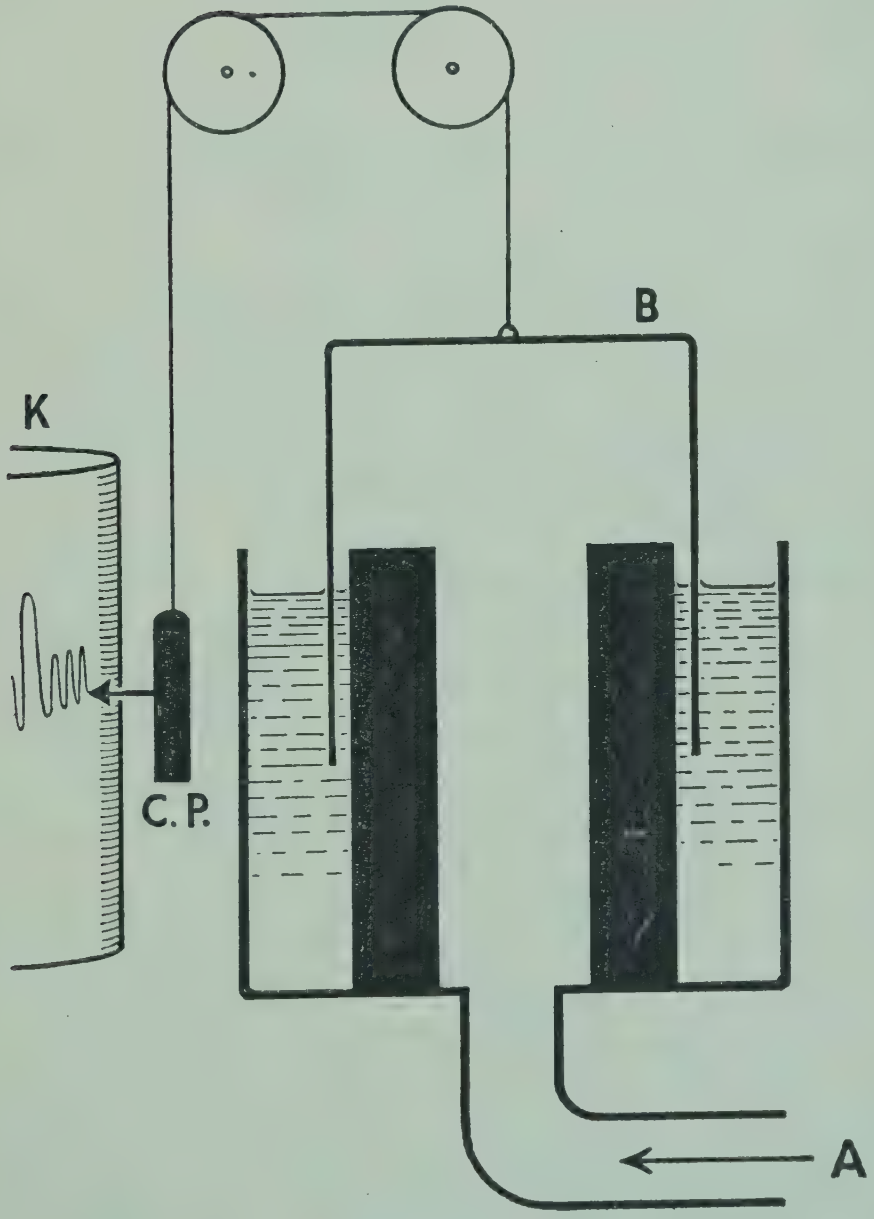


FIG. 32, 8

Spirometer designed for high respiratory rates. The subject breathes into A and raises the bell B and the index attached to the counterpoise C.P. moves down the kymograph K. (L. Bernstein, J. L. D'Silva & D. Mendel (1952), *Thorax*, 7, 255.)

chest and connected by a tube to a tambour writing on a smoked drum (Fig. 32, 7). Quantitative records can be obtained by using a recording *spirometer* (Fig. 32, 8). The respiratory movements may also be registered by recording the changes in intrapleural pressure (Fig. 32, 10), or in experimental animals the movements of the larynx or of a slip of the diaphragm.

TABLE 32, 9

Respiratory Rate per minute at Different Ages with the Basal Oxygen Requirement

Age	Respiratory Rate per minute	Basal Oxygen Requirement ml. per min.
At birth . . .	18–40	32
First year . .	25–35	78
2–4 years . .	20–30	87–100
5–14 years . .	20–25	100–175
Adult . . .	12–20	230

The figures in the last column have been calculated from the data in Table 11, 16.

Respiratory Rate.—The respiratory rate at rest varies considerably from subject to subject but in the adult it is 12 to 20 per minute. Faster rates are normally found in young children (Table 32, 9). The respiratory rate is increased by exercise and by emotion. For this reason a true resting rate is not likely to be obtained if the subject is aware that his breathing is being observed.

Intrapleural Pressure.—The pleural space contains only a very thin film of fluid which makes it very difficult to separate the two pleural layers. The situation is comparable to that in which a wet rubber 'sucker' is stuck to a piece of glass. When the sucker is applied to the glass and released the atmospheric pressure holds it in position. If the film of water between the rubber and glass evaporates, then the sucker is much more easily dislodged. Similarly, in the case of the lungs, the parietal and visceral layers of the pleura are kept pressed together by the atmospheric pressure and remain in firm contact. Alterations in the volume of the lungs during breathing occur because of the movements of the thoracic cage. In the process of expansion the elastic tissue of the lung is stretched and provides a force tending to collapse the lungs—this is the so-called *negative intrapleural pressure.* It is greater, the greater the expansion of the lungs and the stretch of the elastic tissue. The intrapleural pressure can be measured by injecting a *very small* quantity of air between the two layers of pleura and connecting this air pocket to a water manometer. Such a record is shown in the upper portion of Fig. 32, 10. In expiration the intrapleural pressure is about plus 1 to minus 4 cm water; in normal inspiration it is minus 4 to minus 10 cm water and in deep inspiration about minus 40 cm water.

The abdominal pressure, which is most easily measured through a tube passed into the stomach, varies in the opposite direction to that in the thorax; on inspiration, when the intrapleural pressure becomes more negative, the diaphragm descends, so increasing the intra-abdominal pressure.

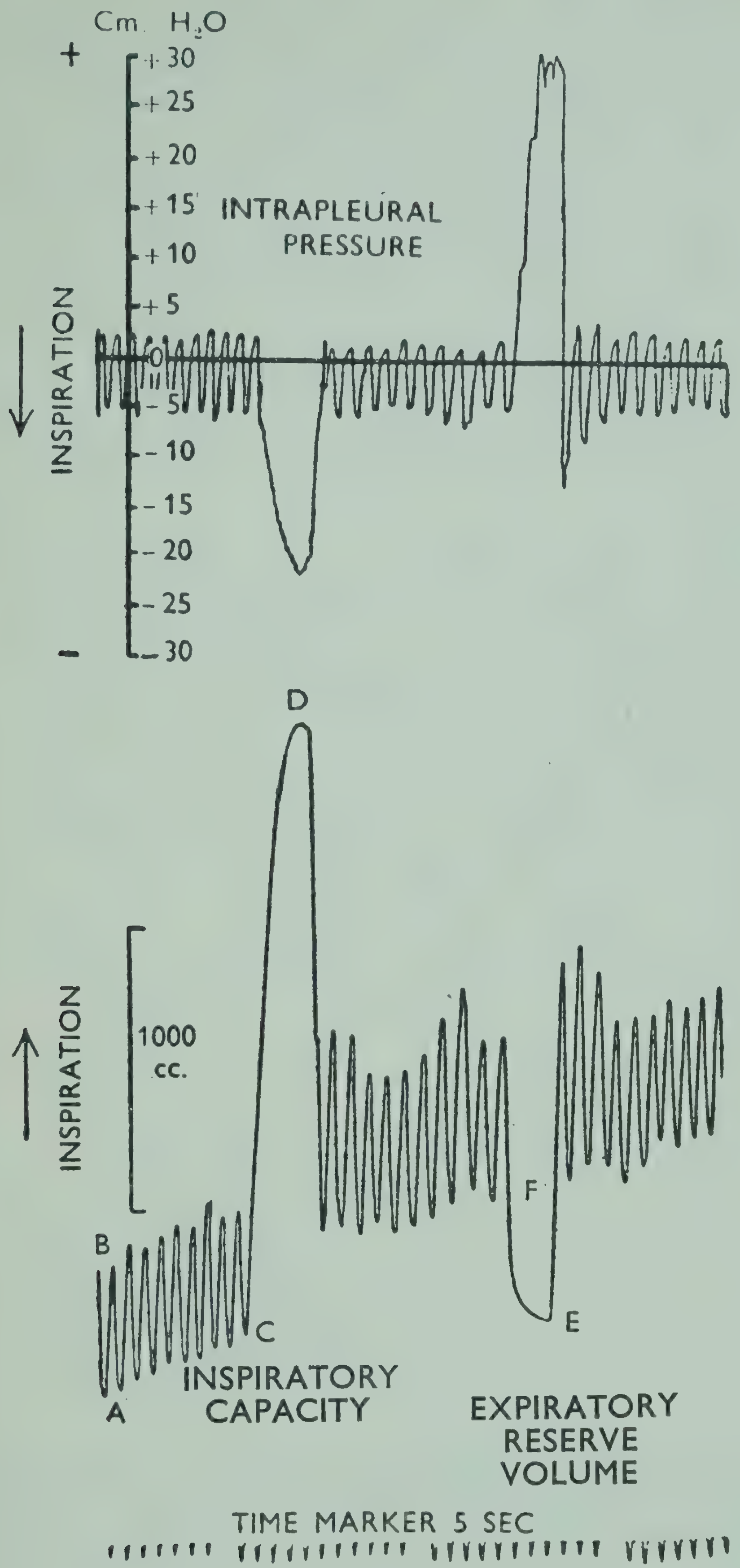


FIG. 32, 10

Simultaneous traces of intrapleural pressure recorded by a water manometer (upper trace) and respiratory record made with a recording spirometer (lower trace). Note on the respiratory trace that AB = tidal volume; CD = inspiratory capacity; EF = expiratory reserve volume; CD + EF = vital capacity (R. V. Christie & C. A. McIntosh (1934), *J. clin. Invest.* **13,** 292).

The elasticity of the lungs can be demonstrated in the following way. If a needle is passed through the chest wall the 'seal' between the two layers of the pleura is broken by the entry of air until the intrapleural pressure is atmospheric and the lung on that side collapses by the retraction of its elastic tissue. This condition, called a *pneumothorax*, is sometimes created therapeutically to render the lung functionless. In new-born animals the lungs fill the thorax completely so that a pneumothorax cannot be produced. It is only later when the thoracic cage has increased in volume at a greater rate than the lungs that this is possible.

Emphysema is a degeneration of the elastic tissue and a distension of parts of the lungs with air. Thus the lungs may not collapse when the pleural cavity is opened; they tend to occupy more space than normal and the residual volume is considerably increased. As the elastic pull of the lungs on the thoracic wall is lost, the chest at the end of a quiet expiration is larger than normal in size. This places the respiratory mechanism at a considerable disadvantage. Many areas in the lungs are inefficiently ventilated, the oxygen saturation of the arterial blood is reduced (anoxic anoxia) and the patient is breathless, especially on exertion.

Bronchial Muscle.—The activity of the bronchial muscle has been studied in the intact animal under artificial respiration provided by a pump with a constant stroke volume. If the bronchi constrict, a manometer connected to the trachea shows an increase of pressure; bronchodilatation is indicated by a fall of tracheal pressure. Excised lungs can be perfused through the trachea with warm Locke's solution which escapes through scratches made on the lung surface. If the bronchi constrict the resistance to perfusion increases and the tracheal pressure rises; bronchial dilatation produces a lowering of the perfusion pressure. In this way it has been found that stimulation of the vagus or injection of acetylcholine produces bronchial constriction while sympathetic stimulation and injection of adrenaline produce bronchodilatation. This last finding is of great practical importance in the treatment of asthma. In this condition the patient has difficulty in breathing and has a sense of suffocation, both of which are produced by a spasm of the bronchial musculature, especially during expiration—an exaggeration of the normal constriction on expiration. The spasm can often be relieved by subcutaneous injection of adrenaline. Stimulation of the vagus or of the sympathetic nerves on one side causes bronchomotor effects on both sides due to crossing over of the fibres in the chest.

The injection of histamine in the guinea-pig causes a marked bronchoconstriction which is probably the cause of death in anaphylactic shock in this animal. Histamine has, however, very little effect on the bronchi in man.

Bronchial tone seems to be regulated by a balance between a large number of effects. Perfusion of the brain with mixed venous blood instead of arterial blood produces a bronchoconstriction due partly to diminished pO_2 and partly to raised pCO_2; these effects occur through variations in vagal tone. Bronchomotor effects in the opposite direction occur reflexly through the aortic and carotid chemoceptors and by direct action of changes in the inspired gas tensions; oxygen lack and carbon dioxide excess by these mechanisms produce bronchodilatation. Lowering of the blood pressure causes bronchodilatation through baroceptors in the aorta and carotid sinuses. Inflation or deflation of the bronchial tree produces a reflex bronchoconstriction (as in coughing, p. 564) by stimulation of slowly adapting receptors, probably smooth muscle spindles, found mainly at bronchial bifurcations where smooth muscle is present. This reflex may account for the changes in calibre of the bronchi in breathing (p. 542).

BREATH SOUNDS

If a stethoscope is applied over the trachea a loud blowing sound can be heard during both inspiration and expiration, with usually a little pause between the inspiratory and expiratory sounds. This *tubular* or *bronchial murmur* can also be heard over the larger bronchi in the chest. These sounds depend to some extent on the narrowing of the air way at the larynx. Over other parts of the healthy chest a very faint *vesicular murmur* is heard beginning with inspiration and continuing into the early part of expiration without any pause. The cause of the vesicular murmur is a matter for speculation. It may be caused by the opening up of the air sacs or by the turbulent passage of air from the terminal bronchioles into the rather wider air passages opening off them; it is possible, however, that it represents the bronchial murmur modified by conduction through the spongy lung tissue. It is difficult to exclude the possibility that the movements of the thoracic cage and muscles contribute to the sound.

The velocity of sound in hydrogen is nearly four times that in air, and consequently if a subject breathes hydrogen all sounds depending on the resonance of gas columns in the lungs become higher in pitch. The tracheal and bronchial sounds do rise in pitch but the vesicular murmur is not altered, so that its causation must be quite different from that of the other sounds.

VOLUME OF AIR BREATHED

The volume of air breathed in and out in a single quiet respiration is about 500 ml. This *resting tidal volume* is measured by causing the subject to breathe from a recording spirometer (Fig. 32, 8). A violent inspiratory effort can take into the lungs about 2500 ml. measured from the resting respiratory level, which by convention is taken as the end of a normal expiration (Fig. 32, 11, and A, C, or F in Fig. 32, 10). This is called the *inspiratory capacity*; it includes, of course, the resting tidal volume as shown in Fig. 32, 10. After a quiet expiration (500 ml.) it is possible by a violent expiratory effort to blow out approximately 1300 ml. of air; this is the *expiratory reserve volume.* Even after the deepest possible expiration the lungs and respiratory passages still contain about 1600 ml. of air; this is the *residual volume.* At the end of a quiet respiration the lungs contain the expiratory reserve volume and the residual volume which together are called the *functional residual capacity* (2900 ml.) (see Fig. 32, 11).

The amount of air breathed in per minute is called the respiratory minute volume or pulmonary ventilation. Since the respiratory rate at rest is about 16 per minute and the tidal volume 500 ml., the minute volume is about 8 litres. In exercise it may go up to 70 litres.

Measurement of Functional Residual Capacity.—This volume can be measured by diluting the gases in the lungs with a known quantity of oxygen and measuring the amount of the dilution. The following method is used clinically. A recording spirometer with a carbon dioxide absorption apparatus (Benedict-Roth apparatus, p. 159) is first filled with oxygen, then at the end of an ordinary expiration the subject is connected by tubing to the spirometer. After a few minutes of quiet breathing the volume of gas in the spirometer is noted and its nitrogen content measured. The nitrogen originally in the subject's lungs plus the nitrogen in the connecting tubing has now been distributed evenly through the gas in the spirometer, the tubing and the lungs, and therefore

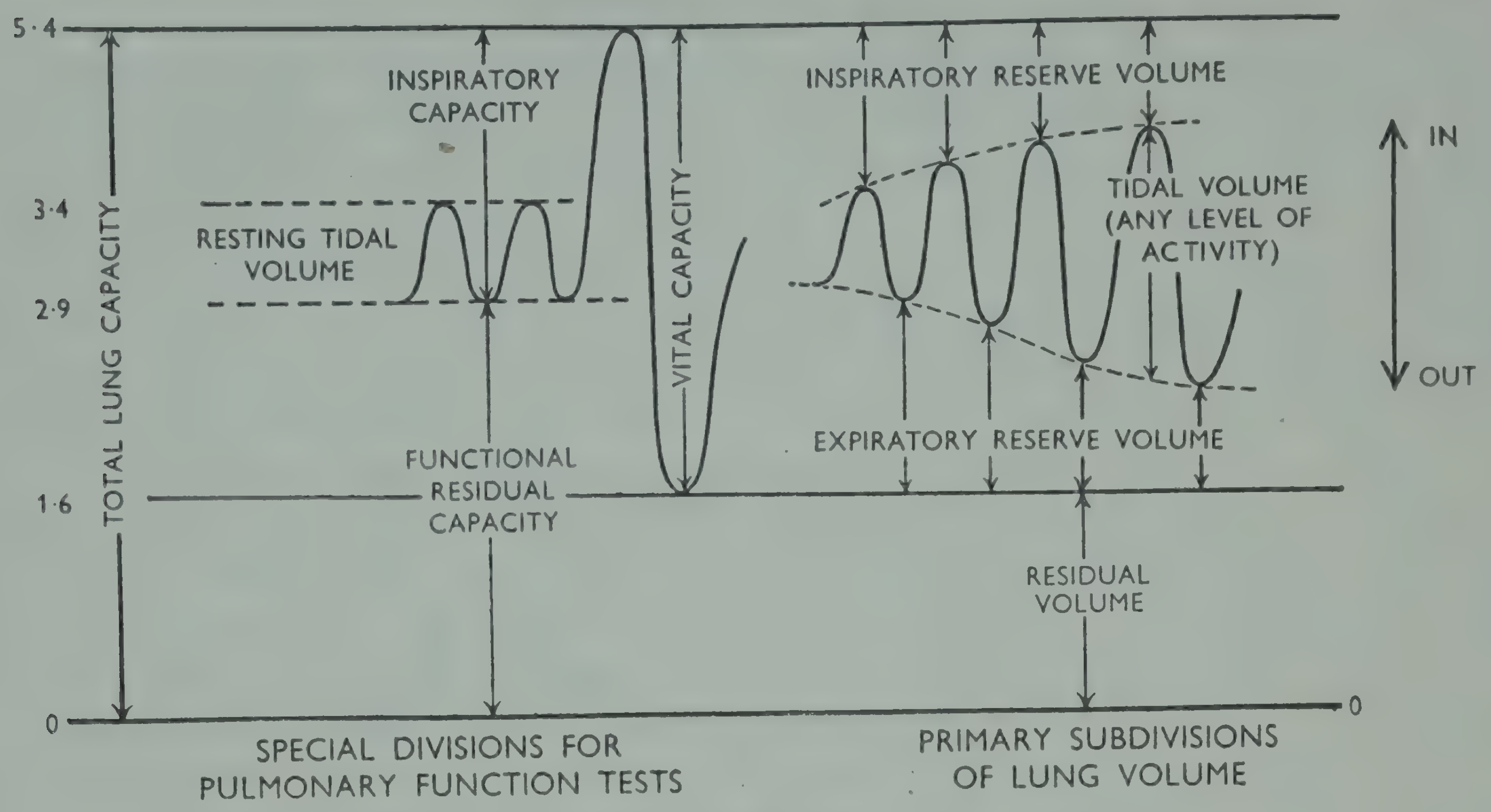

5·4
3·4
2·9
1·6
0
0
TOTAL LUNG CAPACITY
INSPIRATORY CAPACITY
RESTING TIDAL VOLUME
VITAL CAPACITY
FUNCTIONAL RESIDUAL CAPACITY
INSPIRATORY RESERVE VOLUME
TIDAL VOLUME (ANY LEVEL OF ACTIVITY)
IN
OUT
EXPIRATORY RESERVE VOLUME
RESIDUAL VOLUME
SPECIAL DIVISIONS FOR PULMONARY FUNCTION TESTS
PRIMARY SUBDIVISIONS OF LUNG VOLUME

Fig. 32, 11

A representation of a recording obtained by a spirometer (Fig. 32, 8) to illustrate the subdivisions of the lung volume. The figures on the left are the average values in litres for an adult man. (*Fed. Proc.* (1950), **9,** 602.)

$$L\frac{N_a}{100}+T\frac{79}{100}=(L+T+S)\frac{N_s}{100},$$

where L, T and S are the volumes respectively of the lungs, tubing and spirometer at the end of the experiment; N_a is the percentage of nitrogen in the alveoli at the beginning of the experiment, N_s the percentage of nitrogen in the spirometer at the end of the experiment and 79 the percentage of nitrogen in the atmosphere. Since all the other values are known, L can be calculated thus:

$$L=\frac{N_s(T+S)-79T}{N_a-N_s}$$

For clinical purposes a method in which the subject breathes a mixture of hydrogen and oxygen from a spirometer, into which oxygen is run to keep the trace level, is more convenient because the percentage of hydrogen in the system can be easily measured electrically.

Vital Capacity.—If, after the deepest possible inspiration, a subject makes as forcible as possible an expiration into a spirometer we obtain a reading of the *vital capacity*. Using the figures already quoted we find that this is composed of

Inspiratory capacity . . .	2500
Expiratory reserve volume . .	1300
	3800 ml.

The vital capacity clearly depends on the size of the subject but the correlation between vital capacity and stem height, surface area, or any of the usual physical measurements, is not very close, so that accurate prediction of the normal value for the vital capacity of any one subject is not possible. Furthermore the chest expansion as given by a tape-measure during a deep inspiration does not give a reliable indication of the vital capacity.

The actual readings obtained with the spirometer depend somewhat on the position of the subject, as is shown by the following readings obtained in one man:

Standing erect . . .	4300 ml.
Sitting	4200 ml.
Lying supine . . .	3800 ml.
Lying prone . . .	3620 ml.

The lower values obtained on lying are due partly to the pressure of the abdominal viscera on the diaphragm and partly to the accumulation of blood in the lungs. The vital capacity is reduced by conditions which interfere with thoracic movements, and in diseases in which the air in the spongy lung is replaced by fluid or solid material, and when expansion of the lung is impaired by the presence of fluid or air between the layers of the pleura. In cases of heart disease the vital capacity may be reduced 10 to 30 per cent., partly by an accumulation of blood in the engorged vessels of the pulmonary system, and partly by fluid extravasated into the alveolar spaces. When the vital capacity of a patient is reduced in this way to about 40 per cent. of normal, he can no longer perform even the simplest movements without becoming breathless. Such patients are much more breathless lying down than sitting up (*orthopnœa*), one of the reasons for which is the decline in vital capacity in the recumbent position.

The vital capacity (3800 ml.) and the residual volume (1600 ml.) added together give the *total lung capacity* which is 5400 ml. on the average.

Dead Space.—Of the 500 ml. of air taken in at each quiet inspiration only 360 ml. reach the respiratory parts of the lungs. The remainder occupies the mouth, pharynx, trachea and air passages up to the terminal bronchioles (that is the anatomical dead space), and is unchanged in composition because no respiratory exchange takes place. Since the air expired in a single expiration is a mixture of air from the alveoli in which respiratory exchange has occurred, together with the unchanged air, that is atmospheric air from the dead space, the volume of the latter is easily calculated. Thus, if the alveolar air contains 6 per cent. carbon dioxide, while the expired air contains 4 per cent. carbon dioxide, the expired air must be a mixture of two parts alveolar air and one part unchanged inspired air. If the tidal volume is 500 ml. the dead space in this simplified example is 167 ml., which is about the volume of the anatomical dead space.

For strict accuracy all values for the subdivisions of the lung volume and all results of pulmonary function tests should be given at body temperature and saturated with water vapour and not at 0° C as are metabolic data. The values quoted above apply to adult males: the values in women are about 25 per cent. lower.

Respiratory Efficiency Tests

In addition to the measurement of vital capacity just mentioned, several tests of respiratory function are applied to candidates for training in flying. In the *breath-holding test* a clip is applied to the nose and the subject breathes out as deeply as possible, at the same time bending his trunk between his knees, and then breathes in as fully as he can while raising the trunk to the erect position. The normal fit young adult can hold his breath after this procedure for at least one minute.

In the *endurance* or 40 *millimetre test*, after applying a nose-clip, the candidate breathes out and in fully as in the preceding test and then blows out through a mouthpiece connected to a mercury manometer. He adjusts the expiratory effort so that the pressure is maintained at 40 mm Hg. Since the thoracic and abdominal pressures measured during this test are much lower than 40 mm it seems that the high mouth pressure is maintained by closure of the glottis. Thus the usual time of maintenance in fit people, 50 seconds, is not much less than the simple breath-holding time. It is doubtful if there is any correlation between the result of this test and physical fitness.

The same apparatus can be used to find the *maximum expiratory effort*. In fit young men the mercury can be raised steadily without over-swinging till the pressure is about 110 mm Hg.

One of the most effective tests of respiratory function is the *maximum breathing capacity* (expressed in l. per min.), that is the greatest volume of air that can be breathed in a given period of time, usually 15 seconds, by voluntary effort; it is, thus, a measure of maximum ventilation. The maximum breathing capacity varies with age; normal values in young adults are from 180 to 200 l. per minute. Although there is a high degree of correlation between maximum breathing capacity and vital capacity in normal persons, in disease there may be wide discrepancies, for the maximum breathing capacity depends not only on vital capacity but on numerous other factors, such as willingness to co-operate, muscular co-ordination and power, and a fully patent airway, as well as on the viscous properties of the lungs.

Because of the strenuous nature of the test, some investigators prefer to measure, from a kymographic record of the displacement of a spirometer, the *respiratory flow rate* in a single maximum expiration.

Normally, of course, the volume of air breathed per minute (about 8 l.), falls far short of the maximum breathing capacity, and the difference may be regarded as the *breathing reserve*. It has been shown that when the breathing reserve is reduced to some 60 to 70 per cent. of the maximum breathing capacity dyspnœa is present.

FŒTAL AND NEONATAL RESPIRATION

There is no doubt that the fœtus even from the early age of 12 weeks carries out respiratory movements. This has been shown by injecting thorotrast into the amniotic fluid before Cæsarean section. X-ray photographs of the fœtus immediately after birth show the radio-opaque thorium in its lungs and intestinal tract. Since the thorium becomes concentrated in the lungs the water of the amniotic fluid must be absorbed through the lungs into the fœtal circulation. During intrauterine life respiratory movements are intermittent and irregular. At birth, air is breathed instead of fluid and the respirations become deeper and more regular. The extent of the expansion of the lungs is indicated by the change in specific gravity; lungs removed from the body before respiration has begun sink in water (sp. gr. 1·06) whereas ærated lungs float (sp. gr. 0·34).

The respiratory rate at birth is usually about 30 per minute (range 18 to 40); as the average respiratory minute volume is 600 ml. the average tidal volume is about 20 ml. Breathing may be quite irregular in the first few days of extrauterine life but it soon becomes regular. The respiration of the new born infant is stimulated by administration of carbon dioxide just as is the respiration of the adult (see Table 32, 16).

COMPOSITION OF THE RESPIRED AIR

The effect of respiration, so far as oxygen is concerned, is that an adult man at rest acquires about 250 ml. of oxygen per minute. This is the amount of oxygen which must be supplied when a gaseous anæsthetic is administered by any of the modern closed methods in which the patient is entirely cut off from the atmospheric air. At the same time about the same quantity of carbon dioxide must be got rid of.

Expired air can be collected by making the subject breathe through valves into a rubberized canvas bag (Douglas bag) (p. 161). A sample of *alveolar air* can be obtained by the method devised by Haldane and Priestly. A tube 4 ft. (120 cm) long and 1 in. (2·5 cm) in diameter is fastened to a mouthpiece provided with an evacuated gas sampling tube as in Fig. 32, 12. At the end of a normal expiration the subject closes his nose and breathes out as deeply and as quickly as possible through the tube, finally closing the mouthpiece by putting his tongue against it. Immediately, the upper tap of the sampling tube is opened momentarily to draw in a sample of the air last expelled from the lungs. A similar experiment is made in which the subject breathes out as deeply as possible at the end of a normal inspiration. The mean of the values obtained by analysis is taken as the composition of the alveolar air. In this method the dead space air comes first into the long tube but it is washed out by the alveolar air coming later, so that alveolar air only is taken into the sampling tube. This technique has been criticized since, by interrupting the normal rhythm of respiration, it may alter the composition of the alveolar air which is the very thing it is desired to measure. Recent methods attempt to overcome this difficulty by obtaining a small quantity of the gas at the end of each normal expiration over a period of time.

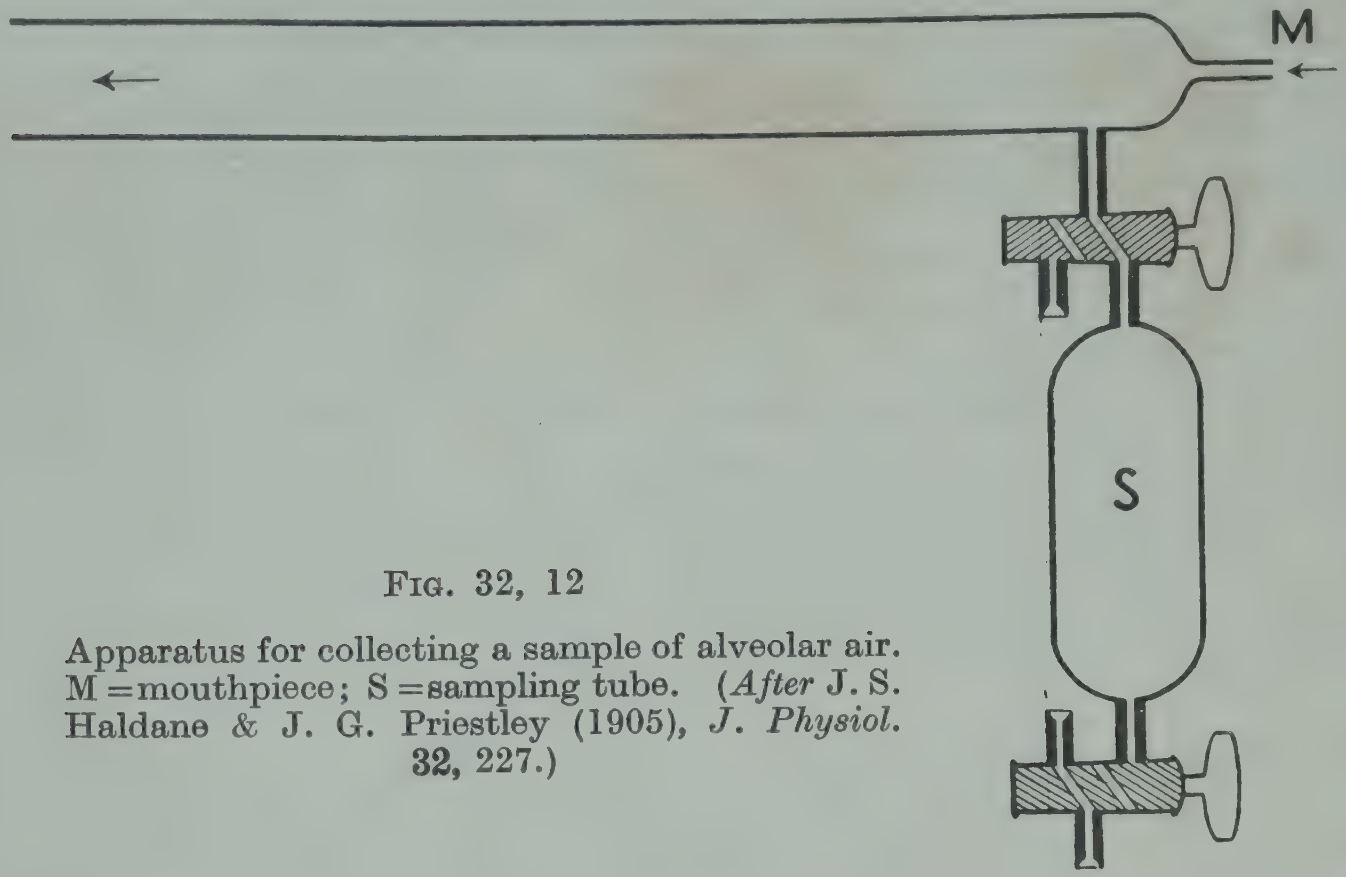


Fig. 32, 12

Apparatus for collecting a sample of alveolar air. M = mouthpiece; S = sampling tube. (*After* J. S. Haldane & J. G. Priestley (1905), *J. Physiol.* **32**, 227.)

Analysis of a sample of air is best carried out by a Haldane gas analysis apparatus in which its volume is first accurately measured. The air is then passed over caustic potash solution to absorb carbon dioxide and the volume is measured once more. Finally the oxygen is absorbed by alkaline pyrogallol solution and the oxygen content is shown by the further reduction in the volume of the sample.

The following table of experimental results shows how the percentage composition of both expired and alveolar air differs from that of inspired (atmospheric) air.

Percentage Composition of Dry Gas	Inspired Air	Expired Air	Alveolar Air
Oxygen . . .	20·95	16·4	13·8
Nitrogen . . .	79·01	79·5	80·7
Carbon dioxide . .	0·04	4·1	5·5

Nitrogen is not given out from the body as the increase in nitrogen content of the expired air might at first sight suggest. Since the respiratory quotient,

$$\frac{\text{Volume of } CO_2 \text{ produced}}{\text{Volume of } O_2 \text{ used}},$$

is usually less than 1 (see p. 154), the volume of carbon dioxide escaping from the blood is less than the volume of oxygen taken into the blood.

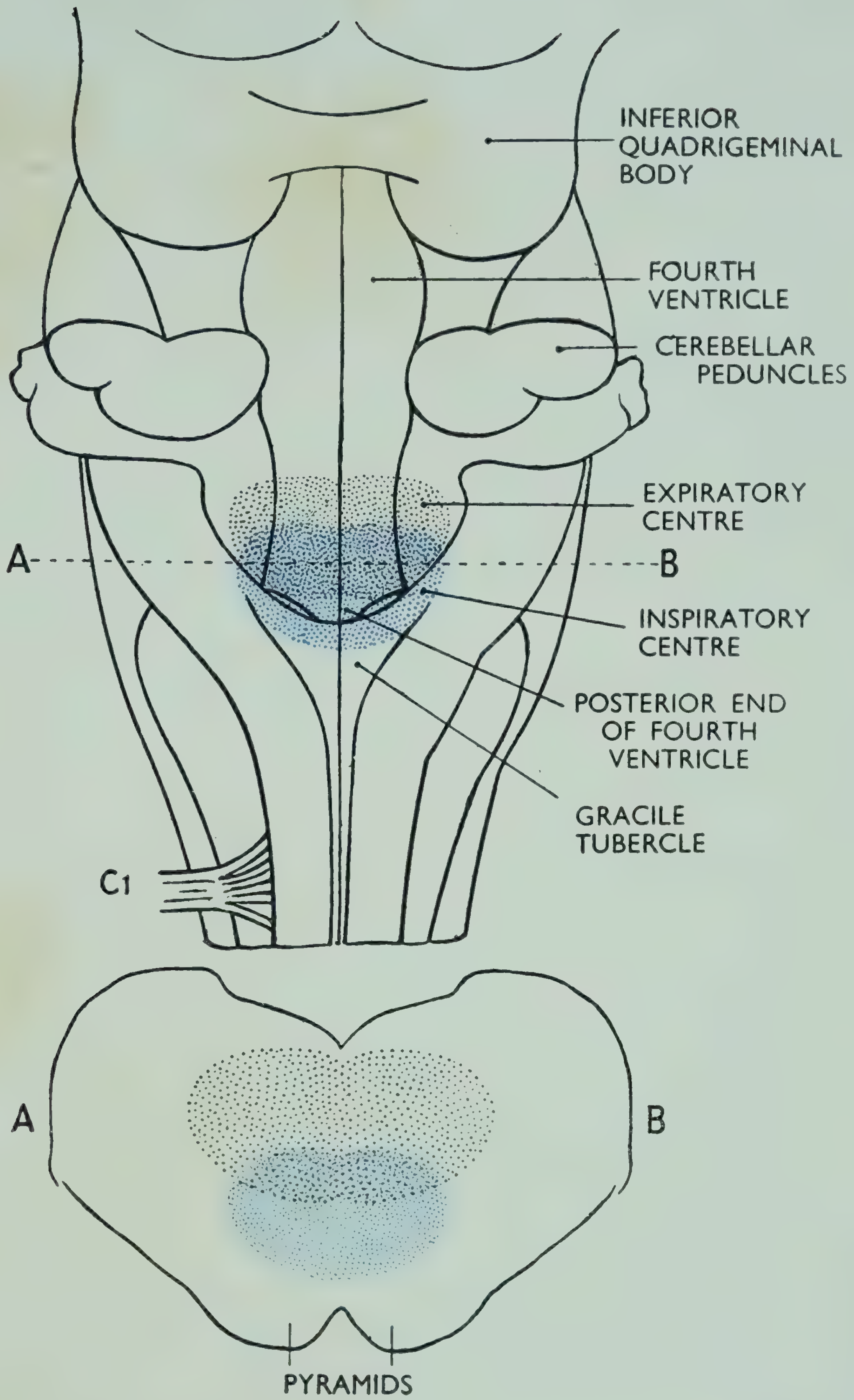

INFERIOR QUADRIGEMINAL BODY
FOURTH VENTRICLE
CEREBELLAR PEDUNCLES
EXPIRATORY CENTRE
A
B
INSPIRATORY CENTRE
POSTERIOR END OF FOURTH VENTRICLE
GRACILE TUBERCLE
C1
A
B
PYRAMIDS

Fig. 32, 13

Diagrams to show the position of the respiratory 'centres' in the reticular formation of the medulla of the cat. The centres lie below the posterior part of the fourth ventricle, as shown in the upper diagram; the depth below the floor of the ventricle is shown in the lower diagram, which is a transverse section through the line AB. The expiratory centre is shown in black stipple, the inspiratory in blue stipple (R. F. Pitts, H. W. Magoun & S. W. Ranson (1939), *Amer. J. Physiol.* **126,** 678).

[*Facing page* 552

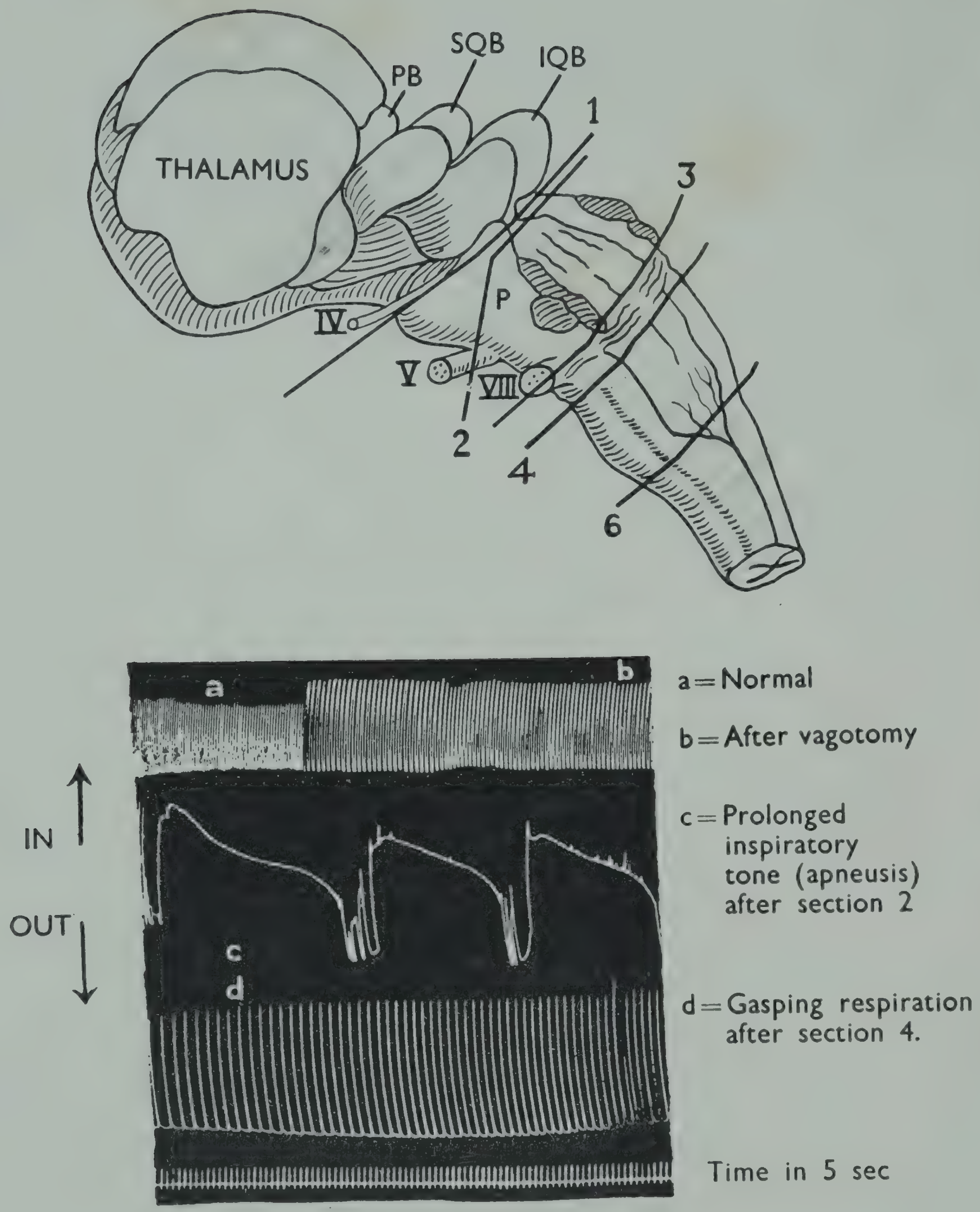

SQB
IQB
PB
1
THALAMUS
3
P
IV
V
VIII
2
4
6
a
b
a = Normal
b = After vagotomy
IN
OUT
c
d
c = Prolonged inspiratory tone (apneusis) after section 2
d = Gasping respiration after section 4.
Time in 5 sec

Fig. 32, 14

Diagram (not to scale) of the brain stem of a cat to show the position of sections which produced the alterations in respiratory rhythm shown in the lower part of the figure. PB, pineal body; SQB and IQB, superior and inferior quadrigeminal bodies; P, pons; IV, V, VIII, cranial nerves (T. Lumsden (1923), *J. Physiol.* **57,** 153, 354; **58,** 81, 111).

Some of the inspired oxygen is apparently 'lost' because it forms compounds, with atoms such as hydrogen, which are not returned as gases to the lungs. Thus the volume of air expired is less than the volume of air inspired although both contain the same absolute amount of nitrogen. The R.Q. in the example given above is 0·87, and this accounts for the differences in the nitrogen percentages.

The alveolar air is not so variable in composition as the tidal air. During quiet breathing the small quantity of air drawn into the alveoli with each breath—360 ml.—is not sufficient to disturb very greatly the composition of the large volume of functional residual air, 2900 ml., but in deep breathing the composition of the expired air approaches more closely that of the alveolar air. It might be thought that such a system of ventilation of the alveoli is inefficient since the alveoli are not swept through with fresh air at each breath. Calculation shows, however, that the known rates of diffusion of oxygen and carbon dioxide are more than sufficient to allow the rapid transfer of carbon dioxide and oxygen through the space occupied by the functional residual air. In any case if the alveoli were swept through with fresh air the partial pressure of carbon dioxide would fall and carbon dioxide would escape too freely from the blood with consequent upset of the pH of the blood. The high partial pressure of carbon dioxide in the alveoli serves to retard the escape of carbon dioxide from the blood.

CENTRAL CONTROL OF RESPIRATION

It is already apparent that breathing is a very complicated act involving the co-ordinated movements of many striated and smooth muscles. The integration of this complex process is carried out by a part of the brain stem called for convenience the *respiratory centre.*

If the upper parts of the brain are cut away down to the mid-brain respiration continues, but section through the posterior part of the medulla brings respiration to a stop. When the spinal cord is cut across in the thoracic region the intercostal muscles innervated from levels below the section are paralysed although the diaphragm, supplied by the phrenic nerves, continues to move. Division above the roots of the phrenic nerves (3rd, 4th and 5th cervical) arrests diaphragmatic movements and the animal dies of asphyxia, although for a time movements of the alæ nasi may be seen. The respiratory centre must, therefore, lie somewhere in the hind-brain and must send impulses out along the phrenic and intercostal nerves.

The centre has been more accurately localized in animals by exploring the medulla with electrodes passed into its substance. Electrical stimulation of an area (Fig. 32, 13) dorsal to the pyramids and deep to the posterior (caudal) end of the fourth ventricle produces a prolonged inspiration; this area can be called the 'inspiratory centre.' The 'expiratory centre' lies immediately dorsal to the 'inspiratory centre' and only a short distance below (that is, ventral to) the floor of the posterior part of the fourth ventricle. Stimulation of this area gives a prolonged contraction of the expiratory muscles. A reciprocal action can occur between the two centres, which are connected by a large number of fibres, so that if the inspiratory centre is stimulated the expiratory centre is inhibited. This description applies to the cat and monkey; in other species the two centres are not so clearly separable.

Lumsden found that, after cutting the vagus nerves, section through 1 (Fig. 32, 14) left respiration unaltered (*b*). Section between 2 and 3 produced prolonged inspiration as at (*c*). Section at 4 produced gasping respiration (*d*). Section at 6 abolished all respiratory movements. These results show that there is a controlling or *pneumotaxic centre* in the upper (cranial) part of the pons which controls the activities of lower (caudal) groups of neurones, that is, those causing prolonged inspiration and gasping. If the vagal fibres which pass mainly to the expiratory centre are left intact after cutting between 2 and 3 (Fig. 32, 14) breathing is still regular. Thus to maintain rhythmic breathing the respiratory centres in the lower part of the medulla must be influenced by either the pneumotaxic centre or by afferent impulses through the vagus nerves (Fig. 32, 15). Recent experiments on the dog have amplified Lumsden's findings. The apneustic state after section at level 2 (Fig. 32, 14) is greatly reduced by denervation of the carotid bodies. Dogs transected at level 4 with the vagus nerves cut can breathe regularly for many hours.

While both the rate and the depth of respiration can to some extent be controlled voluntarily, and even arrested altogether for a short time, the respiratory centre seems to possess a rhythmic activity of its own. This was first clearly demonstrated in the goldfish since after removal of the brain from the body the medulla shows rhythmic electrical variations with the same frequency as the gill movements. An analogous intrinsic rhythm in mammals after complete denervation is abolished by artificial respiration with oxygen and increased if carbon dioxide is added to the inspired air.

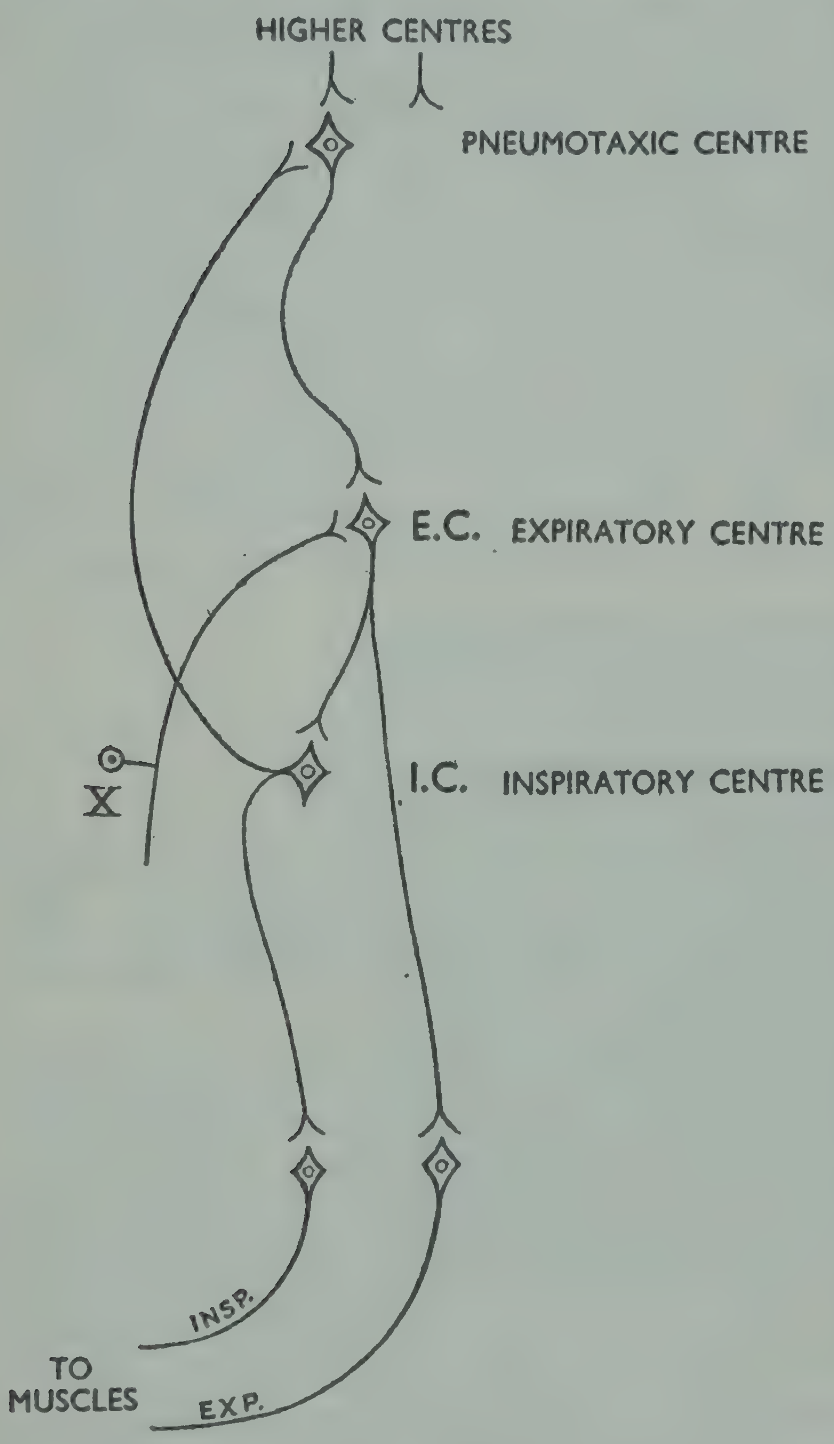


Fig. 32, 15

A simplified scheme of the various parts of the central respiratory connexions. X, vagus nerve. (*After* R. F. Pitts, H. W. Magoun & S. W. Ranson (1939), *Amer. J. Physiol.* **126,** 702).

Stimulation of certain parts of the premotor cortex increases respiratory movements, whereas stimulation of other cortical areas inhibits them. Cooling the medulla slows respiration. If the body temperature of an animal is raised, or if the blood going to the medulla is warmed, respiration becomes faster but not deeper (*tachypnœa*). During sleep the respiratory centre is less active so that the alveolar carbon dioxide may rise above the waking value.

Fibres pass down from the respiratory centre, through the anterior and antero-lateral columns of the cord to the nuclei of the phrenic and intercostal nerves in the anterior horns of the grey matter. Commissural fibres join the nuclei across the spinal cord, since hemisection of the cervical cord does not always cause paralysis of the diaphragm on the same side.

CHEMICAL REGULATION OF RESPIRATION

Asphyxia.—If oxygen is prevented from reaching the blood, the phenomena of asphyxia appear. The circulatory phenomena are described on p. 514. The respiratory events occur in three stages.

(1) The respiratory movements are first increased in depth and rate. The breathing becomes very violent and the expiratory efforts more and more powerful. Consciousness is lost at the end of this stage which is followed by (2) the stage of expiratory convulsions involving practically all the muscles of the body, with stimulation of many central neurones, resulting in vasoconstriction, salivation, contraction of pupils and disturbance of intestinal movements. This is followed quite suddenly by (3) a stage of slow deep inspirations in which most bodily activities are in abeyance and reflexes are absent. Finally the respirations become infrequent and ineffective and the animal dies in a few minutes of acute oxygen want.

Effect of Carbon Dioxide.—The effects of asphyxia on the circulation are produced partly by excess of carbon dioxide and partly by lack of oxygen in the blood and in the tissues, and the same is true of the respiratory phenomena of asphyxia. The effect of carbon dioxide is well illustrated in the experiments of Haldane and Priestley in which the subject breathed air containing varying amounts of carbon dioxide up to 6 per cent. (Table 32, 16). The first effect is an increase in the depth of respiration (*hyperpnœa*) rather than the rate. With low concentrations of carbon dioxide

TABLE 32, 16

*The Effect of increasing the Carbon Dioxide Content of Inspired Air.**

Percentage of CO_2 in inspired air .	0·04	0·79	2·02	3·07	5·14	6·02
Average depth of respiration, ml.	673	739	864	1216	1771	2104
Average frequency per minute .	14	14	15	15	19	27
Ventilation of alveoli, 100 = normal	100	116	153	226	498	857
Percentage of CO_2 in alveolar air .	5·6	5·5	5·6	5·5	6·2	6·6

---

* J. S. Haldane & J. G. Priestley (1905), *J. Physiol.* **32**, 248–249.

the subject may not notice this, but at concentrations above 3 per cent. there is *dyspnœa* which has been described as 'the consciousness of the necessity for an increased respiratory effort.' If air containing between 3 and 5 per cent. carbon dioxide is breathed respiration becomes not only deeper but also more rapid, while at concentrations over 6 per cent. cardiac and respiratory functions are depressed, confusion and headache occur, consciousness is lost and death follows shortly after.

At concentrations of carbon dioxide up to 4 or 5 per cent. the greater depth of respiration has the effect of mixing the alveolar air with a greater amount of inspired air. This 'washes out' the alveoli more thoroughly so that, in spite of the high carbon dioxide content of the inspired air, the alveolar carbon dioxide is kept remarkably constant, at a value which varies from person to person but is usually about 5·5 per cent. If the percentage of carbon dioxide in the inspired air exceeds this value, clearly no physiological adjustment could bring down the alveolar air carbon dioxide concentration to 5·5 per cent.

When air containing excess carbon dioxide is inspired the alveolar carbon dioxide concentration rises and the blood leaving the lungs in the pulmonary veins contains more carbon dioxide than normal. When this blood reaches the medulla the respiratory centre is stimulated, the depth, and if necessary the rate, of breathing is increased and maintained until the arterial carbon dioxide, and therefore the alveolar carbon dioxide, is brought back to the normal tension. Further evidence of this is provided by measuring the carbon dioxide percentages in alveolar air obtained from a subject in a pressure chamber. As the pressure rises the alveolar carbon dioxide percentage falls, the partial pressure of alveolar carbon dioxide remaining constant. The arterial carbon dioxide tension is in this way kept constant, the usual value being approximately 40 mm Hg (Table 32, 17).

TABLE 32, 17

Mean Results for Alveolar Air at Different Pressures *

	Barometric Pressure in mm of Hg	Percentage of CO_2 in dry Alveolar Air	Percentage of O_2 in dry Alveolar Air	Partial pressure of CO_2 in Alveolar Air in mm Hg	Partial pressure of O_2 in Alveolar Air in mm Hg
Top of Ben Nevis	646·5	6·615	13·19	39·7	79·0
Oxford	755	5·95	13·97	42·1	98·9
Bottom of Dolcoath	832	5·29	14·74	41·5	115·7
Brompton Hospital compressed air chamber	1260	3·52	16·79	42·7	203·7

* *After* J. S. Haldane & J. G. Priestley (1905), *J. Physiol.* **32,** 232.

The carbon dioxide tension, pCO_2, in the alveoli is calculated as follows. In alveolar air at sea-level the carbon dioxide content is 5·5 per cent. and the pressure of water-vapour 47 mm Hg. Since the atmospheric pressure is 760 mm Hg, the pCO_2 is equal to 5·5 per cent. of (760 minus 47) mm Hg, that is 39 mm Hg.

Apnœa.—At the end of a normal inspiration the breath can be held for half a minute or longer. The time during which the breath is held is called the period of *voluntary apnœa.* If forced breathing (voluntary hyperpnœa) is previously carried out for a short time the period of apnœa is much longer. This cannot be due to an increased oxygen content of the blood, since it is normally very nearly (95 per cent.) saturated with oxygen. The hyperpnœa flushes out the pulmonary alveoli with fresh air and reduces their carbon dioxide content. Since this increases the gradient between the blood and the alveolar air more carbon dioxide than normal escapes from the blood. When this blood of reduced pCO_2 reaches the respiratory centre it does not possess its normal stimulating effect and the breath is easily held. The pCO_2 of the blood gradually builds up again as carbon dioxide from tissue metabolism enters it and finally, at the 'breaking-point,' when the alveolar air contains about 7 per cent. carbon dioxide, the subject is compelled to breathe once more. The period of voluntary apnœa immediately after exercise is relatively short because the pH of the blood is lowered by its increased content of carbon dioxide and by the appearance in it of organic acid radicles produced by muscle metabolism. If hyperpnœa is carried to excess so much carbon dioxide is blown off that the pH of the blood rises and a state of alkalæmia occurs, with many of the signs of tetany, including spasm of the muscles and hyperexcitability of the peripheral nerves (see Ch. 51). If oxygen is breathed during the preparatory hyperpnœa the apnœa is not greatly prolonged. There is thus no doubt that the carbon dioxide tension of the blood is the main factor determining the length of the apnœic period although when there is oxygen lack the breaking-point is reached somewhat sooner.

Effect of Lack of Oxygen.—The alveolar oxygen tension is not at all rigidly regulated. It depends very largely on the percentage of oxygen in the atmosphere and upon the atmospheric pressure (Fig. 32, 18) but to a minor extent only on the depth of ventilation. At sea-level the percentage of oxygen can be reduced from 21 to 13 without producing any marked alteration of respiration. The effects of lowering the oxygen tension still further can be seen by making a subject breathe from a Douglas bag through a soda-lime container to absorb carbon dioxide. He becomes blue (cyanosis, p. 587) as the oxygen is used up but he may not show any respiratory distress. When the oxygen percentage falls below 13 respiration is stimulated by lack of oxygen and more oxygen is taken into the alveoli. As a result of the deeper breathing carbon dioxide is washed out of the alveoli and respiration becomes depressed for a little, only to become deeper as carbon dioxide accumulates and oxygen diminishes. Thus respiration may increase and diminish in depth periodically and may not be sufficient for adequate saturation with oxygen of the arterial blood In normal circumstances, therefore, respiration is little altered by changes in the oxygen content of the blood though it is easily affected by alterations in the carbon dioxide content. It is important to note that the increase in respiration when there is a gross oxygen lack is a reflex effect mediated by the chemoreceptors of the carotid body (p. 563).

Effect of High Altitude.—Although the percentage of oxygen in the air remains the same as we ascend, the atmospheric pressure and the partial pressure of oxygen pO_2 fall. If the barometric pressure falls from 760 to 470 mm Hg, as it does in ascending from sea-level to 12,000 feet (3660 metres), the pO_2 falls by the same amount as in the experiments just described in which the oxygen in the inspired air at sea-level was reduced from 21 to 13 per cent. Above this height the respiration may be periodic as described in the previous paragraph. The changes in pO_2 and pCO_2 of the alveolar air and in the oxygen saturation of the blood at various heights are given graphically in Fig. 32, 18. Continued residence at a high altitude produces acclimatization. In the initial stages respiration is stimulated probably by the chemoreceptors, so that the alveolar pO_2 is raised and the alveolar pCO_2 is reduced; in consequence the blood pH rises. As acclimatization continues the kidney excretes base (see p. 642)

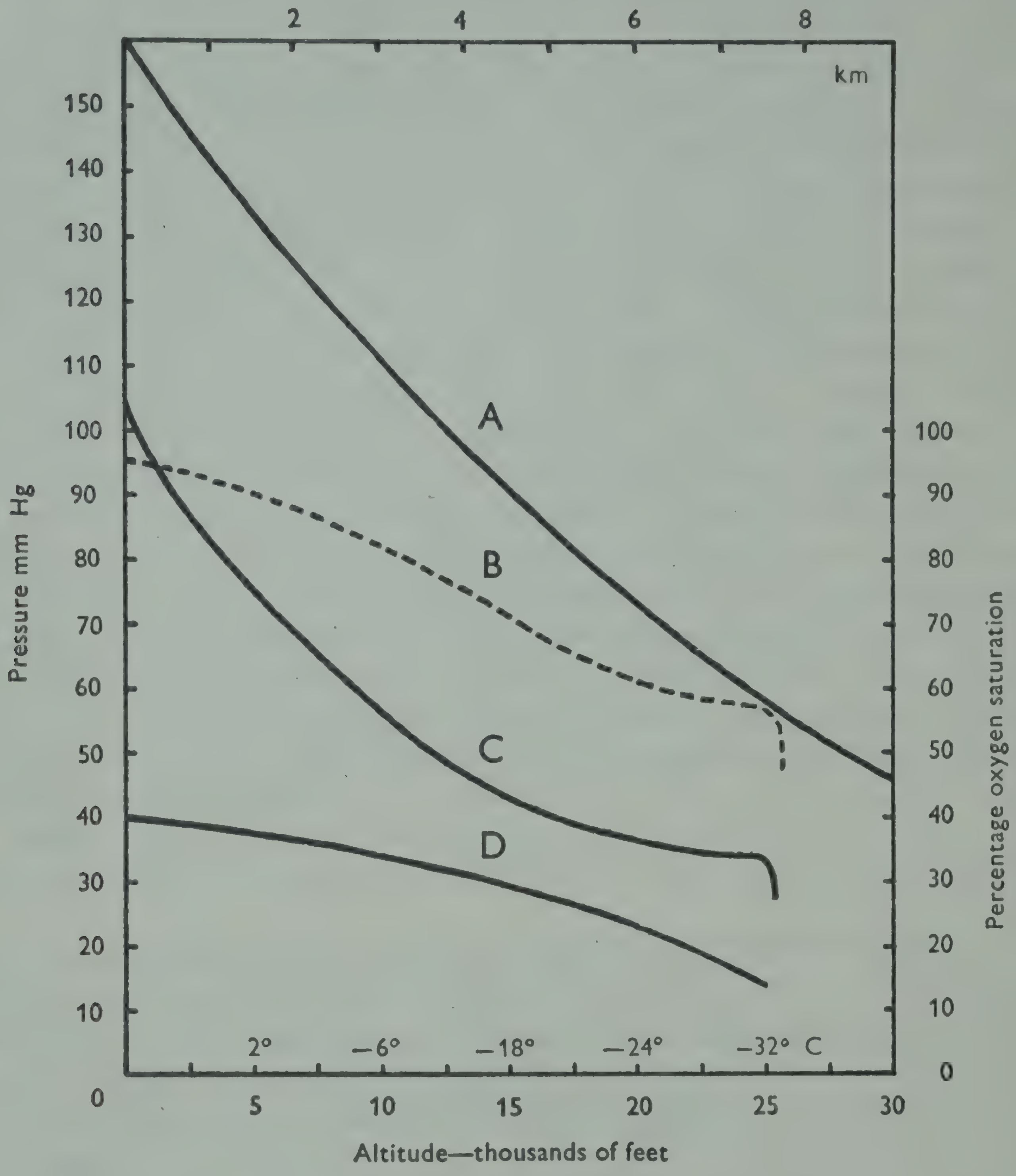


FIG. 32, 18

Curves showing, (A) the partial pressure of the atmospheric oxygen, (B) the oxygen saturation of the blood passing through the lungs, (C) the partial pressure of the oxygen in the lungs, and (D) the partial pressure of the carbon dioxide in the lungs at all altitudes between sea-level and 30,000 feet. The mean air temperature at various heights is given above the abscissa. (*After* M. C. Grow & H. G. Armstrong (1941), *Fit to Fly*, p. 199. New York: D. Appleton Century Co.).

and brings the pH back to its normal value. The respiration is then once more under the control, almost solely, of the respiratory centre and the chemoreceptors are not involved. In the early days of acclimatization mountain sickness may occur; the symptoms, which are due to cerebral anoxia, are weakness, nausea, loss of appetite, vomiting, headache and sleeplessness. The increase in the hæmoglobin content of the blood at high altitudes has already been referred to (p. 386). It is likely also that the muscle hæmoglobin increases under these circumstances. These adaptive changes have enabled fit men to carry out for a short period moderately hard physical work on mountains at 20,000 ft. (6·1 km). The highest permanent human habitation is in the Andes at 17,500 ft. (5·3 km).

When the change from sea-level to high altitudes is rapid as in modern fighter aircraft there is no time for acclimatization and if oxygen is not breathed pilots show signs of cerebral anoxia although they themselves are quite unaware that their behaviour is in any way abnormal. These signs are defects in judgement and reasoning with a spurious feeling of confidence, reduction in visual acuity, weakness of arms and legs with an unsteady gait and twitching, hilarity or pugnacity going on imperceptibly to unconsciousness.

Cheyne-Stokes Respiration.—The commonest type of periodic breathing, is the occurrence of a series of respirations which gradually increase in depth from a barely perceptible movement to a maximum and then progressively decline to terminate in a period of apnœa lasting from a few seconds to as long as a minute. During the apnœic phase the patient may become confused or even unconscious but breathing gradually begins again and the whole cycle is repeated. Cheyne-Stokes breathing occurs in healthy people at high altitudes and occasionally in deep sleep; it is quite common in healthy infants. When it occurs in an adult during waking hours it is of serious significance since it betokens depression of the respiratory centre. The normal stimulus to respiration is insufficient to activate the insensitive centre so that breathing stops. The increasing hypoxia together with accumulation of carbon dioxide ultimately stimulate the centre and breathing begins again. Reoxygenation of the respiratory centre results and the pCO_2 of the arterial blood is reduced by the vigorous respiratory effort to a level which no longer stimulates the respiratory centre. Apnœa then recurs. It is probable that this simple explanation is not always adequate to account for the phenomenon. A reduced concentration of tissue bicarbonate is probably a contributory factor as well as oversensitivity of congested lungs to inflation.

Cheyne-Stokes respiration may be relieved by the inhalation either of carbon dioxide or of oxygen. It is found in patients with chronic heart and kidney disease.

Effect of Increased Oxygen Supply.—It has been suspected for a long time that it is dangerous to breathe pure oxygen but the first satisfactory experiments were made during the last war to investigate casualties in divers carrying self-contained oxygen apparatus. Fit men were exposed in a 'dry' pressure tank to oxygen at 3·7 atmospheres (corresponding to a depth of 27 metres). After exposure times from 6 to 96 minutes severe symptoms such as nausea, vertigo, severe twitching of muscles or even convulsions occurred which necessitated the termination of the experiment. For some unknown reason the oxygen tolerance at the same pressure under water was greatly reduced. Thus in the 'dry' experiments 50 per cent. of the subjects were eliminated by marked symptoms in 20 minutes whereas in the 'wet' experiments 50 per cent. were eliminated in 10 minutes. There is a wide range in individual tolerance to oxygen and in any one individual tolerance varies greatly and unpredictably from day to day. Bradycardia is a well-known effect of exposure to oxygen. Oxygen tolerance is markedly reduced by work but the reason for this is quite unknown. Diving on pure oxygen to 7·6 metres for two hours is quite safe but at greater depths diving on pure oxygen is a hazardous gamble. At 10 metres convulsions frequently occur although at this pressure the oxygen dissolved in the plasma is insufficient even for basal metabolism, and the hæmoglobin must still be used for oxygen transport. Carbon dioxide does not accumulate in the tissues. It is to be noted that these experiments had to be terminated chiefly on account of nervous symptoms and there was no lung irritation or damage such as has been seen in small experimental animals exposed to oxygen. A clue to this may be provided by the results of exposure of rat brain slices to high oxygen pressure; respiration is irreversibly poisoned probably by inhibition of the mechanism for the oxidation of pyruvic acid which is, of course, concerned with the metabolism of carbohydrate (p. 316).

Newborn infants are peculiarly susceptible to oxygen poisoning; administration of oxygen for resuscitation often results in retrolental fibroplasia and blindness. Oxygen administered experimentally to young animals produces first a constriction and obliteration and later a proliferation of the retinal vessels.

DECOMPRESSION SICKNESS.—This condition was first noted in divers and caisson workers who had been brought from a pressure of several atmospheres too quickly to the atmospheric pressure at the surface of the water. Either immediately or after a latent period which may be several hours the victim experiences pain in the limbs (the 'bends') or a choking sensation and may in more severe cases exhibit various forms of paralysis. There is now no doubt that the symptoms are due to the presence of gas bubbles either in the blood vessels or in the nervous system.

The tissues of a man working under increased pressure gradually come into equilibrium with the increased partial pressure of nitrogen and oxygen. If he is suddenly decompressed these gases form bubbles which, as they expand, tear up tissues or block blood vessels. The oxygen is quickly used up and disappears but the bubbles of nitrogen remain. Since nitrogen is five times more soluble in lipoid material than in water, fatty tissues take up relatively more nitrogen during compression and give up correspondingly more during decompression. For this reason particularly large bubbles are formed in the central nervous system and give rise to quite irregular and unpredictable anæsthesia and paralysis.

Decompression sickness can be avoided to a large extent by slow decompression. If a diver suffers from bends on arrival at the surface he should be sent down again and slowly decompressed. A very sudden rise in altitude as may occur in fighter aircraft can also produce similar but not such severe disturbances as those seen in divers.

STIMULATION OF CELLS OF RESPIRATORY CENTRE.—Since a rise of pCO_2 in the blood and an increase of acids in the blood both stimulate the medullary respiratory centre, while both cause a fall of pH, it is pertinent to ask whether the effect of carbon dioxide is a specific property or whether it is due simply to the alteration of pH. Nielsen found that by inhaling a gas mixture containing excess carbon dioxide ventilation was increased by 10 litres per minute when the arterial pH fell by 0·03 units. If the same subject received a long-continued injection of ammonium chloride sufficient to bring about a pH drop of 0·08, his respiratory ventilation increased by only 0·7 litre per minute. In cats, direct injection of bicarbonate buffers into the respiratory centre stimulates respiration much more effectively than injection of lactic acid or hydrochloric acid. It is likely therefore that carbon dioxide has a specific effect on cells in the respiratory centre. These cells lie among the cells of the reticular formation of the medulla which are responsible for the respiratory reflexes (Fig. 32, 13). The cells sensitive to changes in pCO_2 in the blood may be interneurones of the respiratory reflex arcs but, more probably, they may be on side branches converging on these arcs. The latter view is supported by the finding that chloralose anæsthesia reduces the stimulating effect of carbon dioxide on respiration without reducing the stimulating effect of afferent impulses.

Lack of oxygen has a depressant action on the cells of the respiratory centre but in the intact animal inhalation of a gas with a low oxygen content results in increased ventilation. The stimulant effect of oxygen lack on respiration is a reflex effect through the chemoreceptors of the carotid body (p. 563).

NERVOUS REGULATION OF RESPIRATION

There is no doubt that the most important factor in the regulation of respiration is the carbon dioxide tension of the blood passing through the respiratory centre. The activity of the respiratory centre, however, is modified by afferent impulses which are continuously arriving at the medulla in the vagus nerves and in the nerves from the carotid body and carotid sinus; the centre is also influenced, though to a less extent, by impulses in other sensory nerves such as the superior laryngeal. Adrian has recorded afferent impulses in single fibres of the vagus nerve and has found that a stream of impulses passes up the vagus during inspiration (Fig. 32, 19) or artificial distension of the lungs. The stretch receptors cannot be in the visceral pleura since the impulses are not affected by its removal. Since drugs which increase the bronchial tone increase the rate of discharge while drugs which produce a bronchodilatation decrease the discharge the receptors are probably in the bronchioles. The frequency of the sensory impulses rises during inspiration when the lungs are being stretched and falls during expiration when the stretch of the lung tissue is diminishing. Inflation of the lungs in an anæsthetized animal is followed by arrest of respiration. Electrical stimulation of the central end of the cut vagus nerve produces inhibition of respiration which is often followed by a strong contraction of the expiratory muscles (Fig. 28, 34). The exact effect depends on the strength of the stimulus.

Powerful deflation causes a stream of impulses to pass up the vagus and produces a strong inspiratory movement. This reflex, however, plays little part in the control of respiration in ordinary circumstances.

Hering-Breuer Reflex.—In normal breathing the sequence of events is as follows. Inspiration stretches the lungs, afferent impulses increasing in frequency travel along the vagus to the respiratory centre, inspiration is brought to an end and expiration begins, and the frequency of impulses in the vagus gradually falls until inspiration again occurs (Fig. 32, 19). This mechanism is spoken of as the Hering-Breuer reflex. After section or cooling of the vagus nerves the reflex is abolished, inspiration is prolonged and the respiratory rate is reduced (compare *a* and *b*, Fig. 32, 14). It is clear, therefore, that the vagus must convey afferent impulses which shorten inspiration. The respiratory rhythm is probably regulated by the Hering-Breuer reflex while the degree of pulmonary inflation and deflation, that is to say the amount of the tidal volume, is regulated by the carbon dioxide tension in the blood. This view is supported by the finding that double vagotomy does not alter the pulmonary ventilation and by the fact that administration of air with an excess of carbon dioxide, sufficient to cause an increase of respiratory rate and depth in the normal animal, may cause only an increase in depth in a vagotomized animal.

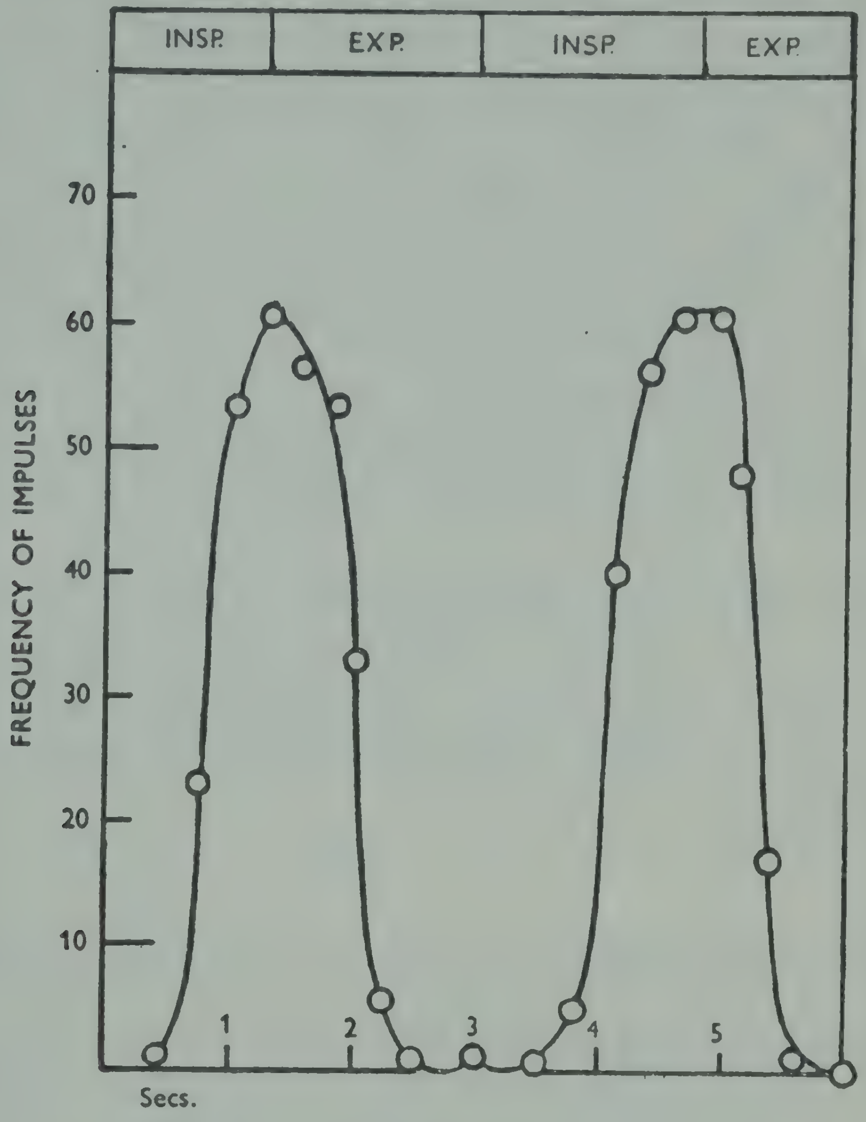


FIG. 32, 19

Frequency of afferent impulses in a single fibre of the vagus of a decerebrate cat during normal breathing. The frequency of nerve impulses at the end of expiration need not necessarily be zero. For example another fibre discharged at the rate of 50 impulses per second at the end of expiration and at 80 impulses per second at the height of a normal inspiration. (*After* E. D. Adrian (1933), *J. Physiol.* **79**, 337.)

Although the dyspnœa of patients with severe congestive cardiac failure may be due in part to stimulation of the respiratory centre by increased carbon dioxide tension of the arterial blood and in part to the effect of oxygen want on the chemoreceptors in the carotid and aortic bodies (see following paragraph) there are many patients with cardiac failure less severe in degree in whom the blood carbon dioxide tension is within normal limits and whose arterial oxygen tension is also normal. In such patients there is considerable engorgement of, and a generalized increase of pressure within, the pulmonary blood vessels, with diminished distensibility of the lung tissue and reduction of its elastic recoil. In these circumstances expiration is no longer a passive act but requires muscular effort so that the intrapleural pressure may actually become positive. In animals increased pressure in the pulmonary capillaries, brought about by temporary occlusion of the left atrio-ventricular junction, is associated with rapid shallow breathing, which is abolished by section of the vagus nerves. The tachypnœa of cardiac failure was at first considered to be a reflex phenomenon due to an increased sensitivity of the Hering-Breuer reflex mechanism but experiment has shown that pulmonary congestion does not sensitize the pulmonary stretch receptors. It is likely that quite a distinct set of receptors situated near the alveoli, which send small diameter fibres into the vagus, is responsible. In congestive cardiac failure the distension of the smaller pulmonary vessels, the presence of pulmonary œdema and the reduced distensibility of the lung tissue all tend to diminish the vital capacity. However, in the less severe case the vital capacity is greater than the tidal volume that the patient requires during exertion and the reduction in the vital capacity by itself cannot be the cause of the dyspnœa.

Carotid Body.—The statement that oxygen lack sensitizes the respiratory centre to carbon dioxide fits the facts so far as they can be observed in intact animals. However, if an animal in which the vagus nerves have been sectioned, and all nerves to the carotid sinuses and the carotid bodies have been cut, is made to breathe air with a much reduced oxygen content paralysis of respiration rapidly occurs. In the intact animal lack of oxygen stimulates the carotid and aortic chemoreceptors and thus reflexly stimulates respiration. These reflexes are usually powerful enough to override the depressant action of oxygen lack on the respiratory centre itself.

The effect of increase or decrease in the pCO_2 of the blood passing through the carotid body is the same as the effect produced on the respiratory centre itself, namely, hyperpnœa or depression of respiratory activity respectively.

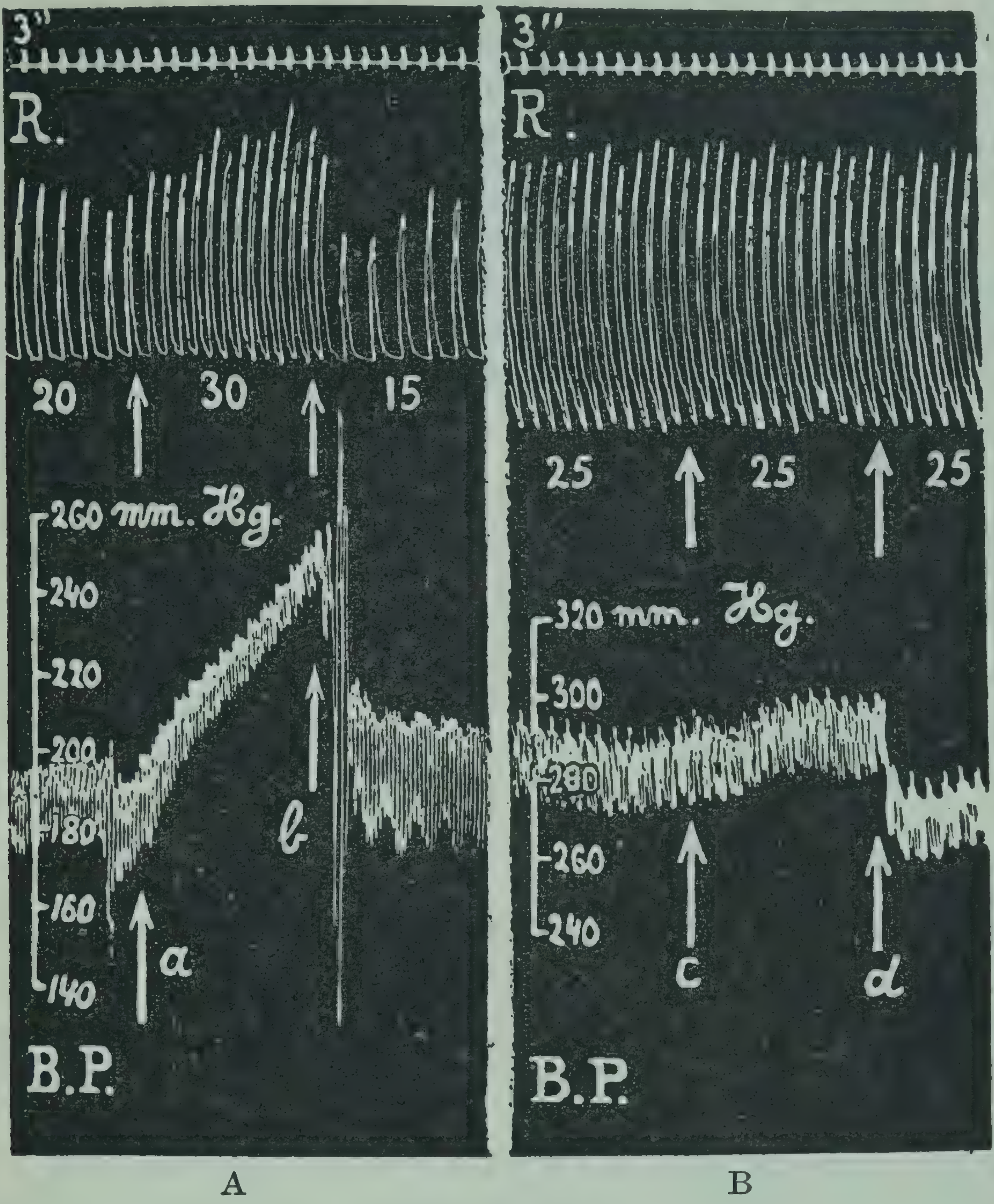
The chemoreceptors are in the carotid bodies at the bifurcation of the common carotid artery and in the aortic bodies which lie between the ascending aorta and the pulmonary artery. Histological preparations of the carotid body show very numerous nerve fibres, some of which supply clumps of large cells and others which arise in the intima of the small vessels. Since the carotid sinus and the carotid body lie very close to one another it has not been easy to decide on their separate functions. The carotid artery and its bifurcation can be perfused artificially and the effect of variations in carbon dioxide and oxygen tension demonstrated. In some animals the artery to the carotid body can be tied off. Since the chemical effects then disappear there is no doubt that the carotid body and not the carotid sinus is responsive to changes in the chemical constitution of the blood. When the aortic bodies and carotid bodies are denervated by cutting the vagi and the carotid sinus nerves the reflex response to chemical changes in the blood disappears; the central effects already described on p. 555 are of course still obtained.

Carotid Sinus.—A rise of blood pressure tends to depress, and a fall to stimulate, respiration. These effects are initiated by the pressure receptors (sometimes called baroceptors) in the aorta and carotid sinuses. When a large dose of adrenaline is injected into an intact animal there is, along with the rise of blood pressure, a cessation of respiration, usually termed 'adrenaline apnœa.' This is a sinus and an aortic reflex and not a direct effect of adrenaline on the respiratory centre, because it is abolished after the vagus nerves and sinus nerves are cut, or if a compensator is included in the circulation to prevent the rise of blood pressure after adrenaline is injected. If the two common carotid arteries are clamped, the pressure within the carotid sinuses falls temporarily to zero; respiration is stimulated and the blood pressure rises (Fig. 32, 20, A). After denervation of both carotid sinuses this effect is absent (Fig. 32, 20, B). Clamping the carotid arteries reduces the number of impulses from the baroceptors but, since it makes the carotid bodies anoxic, it increases the discharge from the chemoceptors. The impulses from the chemoceptors play an appreciable part in the raising of the blood pressure after carotid clamping.

It may be doubted however, if such violent changes of pressure in the carotid sinus are within physiological limits. Both adrenaline and noradrenaline given intravenously in man at the rate of 10 to 20 μg. per minute have a stimulating effect on respiration. Also in the hypertensive crises produced by sudden discharge of these two substances from a medullary tumour the breathing is deep and rapid.

Other Reflex Effects arising outside the Lungs.—A rise of venous pressure in the great veins near the heart, produced for example by rapid injection of fluid, or occurring in cardiac failure, stimulates respiration. Swallowing, which usually occurs during expiration, is attended by a reflex inhibition of respiration (p. 225) via the glossopharyngeal nerve. Stimulation of the superior laryngeal nerve by a foreign body in the larynx causes coughing, a series of forced expirations with the glottis nearly closed. A cough may be useful in bringing mucus or foreign bodies up the bronchi and trachea. Sneezing is a somewhat similar reflex arising from irritation of the nasal mucosa through the nasal branch of the fifth nerve; the expiratory blast passing through the nose may propel droplets up to 20 feet. Hiccough is a spasmodic contraction of the diaphragm, in the course of which the glottis closes suddenly and further entrance of air into the chest is prevented. The sudden arrest of the column of entering air produces the characteristic sound and sensation.

The cough reflexes in man cannot be easily analysed and the results given below of the experiments in the cat may only, in view of the great species variation in these reflexes, be taken as a first approximation to the human conditions. Coughing can be brought about by mechanical irritation of the larynx or trachea but not of the main bronchi. Such irritation also produces a bronchoconstriction. The reflex is not obtained if the mucosa is first sprayed with a local anæsthetic. The receptors are rapidly adapting and are probably knob-like subepithelial endings derived from large afferent fibres. Inhalation of sulphur dioxide or other noxious gases also causes coughing but in this case the receptors must be situated more deeply in the lungs since the reflex is not abolished by anæsthetizing the trachea. The afferent pathways for both reflexes are mainly in the vagus but the sympathetic pathway may also be involved to a small extent.



Fig, 32, 20

Dog 10·5 kg. anæsthetized with chloralosane. Depressor nerves cut. R, respiratory movements. B.P. femoral blood pressure recorded with a mercury manometer. ↑ a, common carotid arteries clamped. This was followed by increased rate and depth of respiration with acceleration of the heart and rise of B.P. ↑ b, common carotid arteries opened. This was followed by a return to original conditions. The two carotid sinus nerves were cut between A and B. This was followed by a rise of B.P. with acceleration of the heart and increase in rate and depth of respiration. ↑ c, common carotid arteries closed. ↑ d, common carotid arteries opened. Only very slight mechanical effects on the B.P. (C. Heymans & J. J. Bouckaert (1930), *J. Physiol.* **69**, 261).

Effect of Exercise on Respiration.—In exercise the metabolic rate rises and the production of carbon dioxide goes up, so that the respiratory centre is stimulated and the rate and depth of respiration are increased accordingly. Very often this hyperpnœa of effort, like the increase in heart rate, begins before the exercise, and for a short time the ventilation may be greater than can be accounted for by the increase in metabolism. This can only be explained as a cortical influence on respiration. Further details are given in the section dealing with exercise on p. 793.

ARTIFICIAL RESPIRATION

There are many occasions when life may be saved by prompt administration of artificial respiration. The main emergencies in which it is required are carbon monoxide poisoning, drowning and electric shock. Probably three or four thousand of such emergencies occur in the United Kingdom per annum. If anterior poliomyelitis (infantile paralysis) affects the motor cells of the diaphragm, prolonged artificial respiration may have to be applied by special apparatus. Artificial respiration is often required during the administration of an anæsthetic, especially if curare is administered, since this drug paralyses all the skeletal muscles, including the diaphragm, by preventing the motor impulses from affecting them. It is important that the physiological principles underlying the methods should be understood so that they may be properly applied.

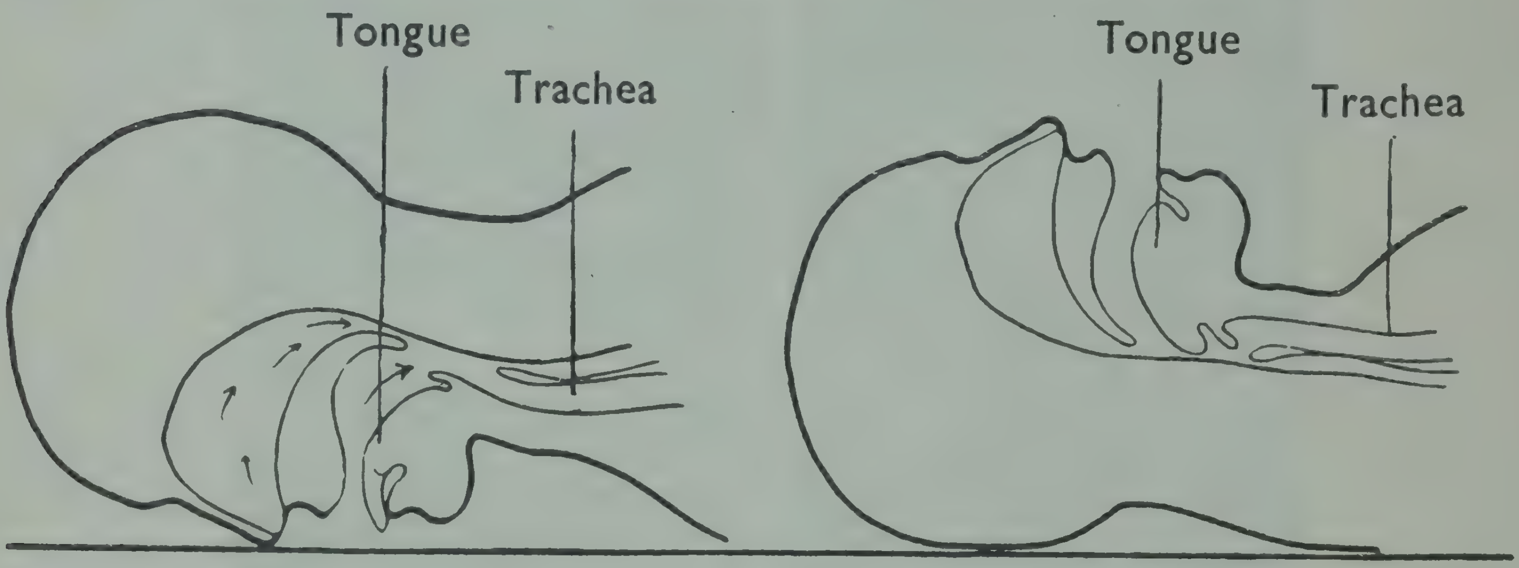


FIG. 32, 21

Effect of the prone position in opening the air passages (Cecil K. Drinker (1938), *Carbon Monoxide Asphyxia*, p. 169. Oxford University Press).

In the emergencies mentioned above respiratory failure occurs first and heart failure only secondarily. If the activity of the medulla oblongata is depressed respiration becomes sluggish and may cease; the heart, on the other hand, continues to beat independently of the central nervous system, always provided that it receives an adequate supply of oxygen. Thus if oxygen can be got into the patient's blood promptly enough by artificial respiration the circulation will be maintained and the respiratory centre has then time to recover its rhythmic activity and restore natural respiration.

To be successful, artificial respiration must be applied promptly because after the cessation of respiration the heart weakens and fairly soon comes to a standstill. An equally important reason for prompt action is that prolonged asphyxia does irreparable damage to the cortical cells of the brain. The oxygen content of the blood falls from 20 to 15 ml. per 100 ml. blood during the first few minutes of submersion and falls to 2 or 3 ml. after six minutes and to zero in eight minutes. If more than eight minutes elapse after complete cessation of respiration there is no possibility of resuscitating the patient completely. Thus even a short delay in beginning artificial respiration may be disastrous; no time should be wasted in loosening clothing or clearing out foreign bodies from the mouth or in moving the victim to a more convenient situation. These matters can be attended to later.

Many methods of artificial respiration have been invented but only a few are of practical value as first-aid measures. The essentials of a good first-aid method are: (1) it should produce adequate ventilation of the lungs; (2) if any apparatus is needed it must be of the simplest kind which should be readily available anywhere; (3) no damage shold be done to the patient by violent handling; (4) the strain on the operator should be a minimum so that he can apply the method for long periods; (5) the position of the patient should be such that the air way is kept open and drainage from the lungs facilitated. The prone position is better than the supine, where the tongue is apt to fall back and block the air way (Fig. 32, 21).

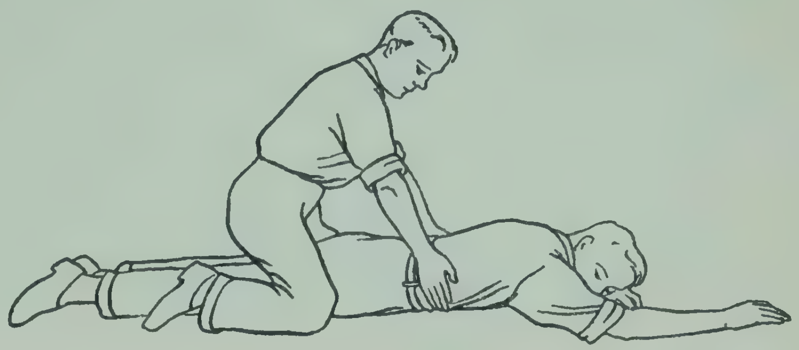


FIG. 32, 22A

Position in which patient should always be placed and kept until conscious; also first position for operator starting artificial respiration by the prone-pressure method.

Prone-pressure Method.—This procedure was first described by Schäfer in 1903 and has undoubtedly saved many lives.

As it is still the method best known to the older teachers of first-aid workers brief instructions are given here. Put the patient in the position shown in Fig. 32, 22A so that the mouth and nose are free for breathing. Kneel, straddling the patient's thighs, and place the palms of the hands on the small of the back, the little finger being on the lowest rib. With arms held straight swing forward slowly so that the weight of your body is gradually brought to bear upon the patient (Fig. 32, 22B). After two seconds swing backwards to the first position for two seconds. Repeat this double movement of compression and release twelve to fifteen times a minute.

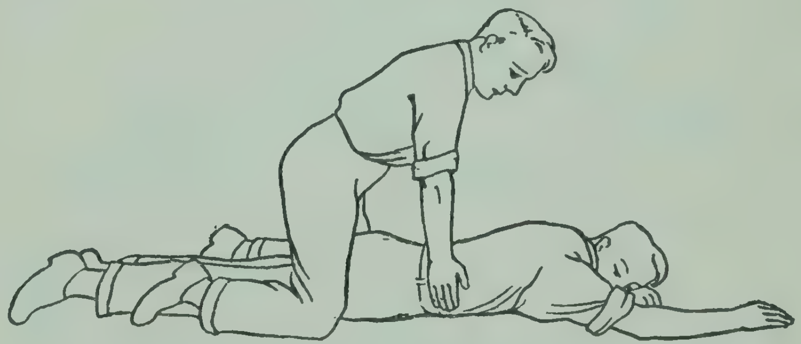


FIG. 32, 22B

Position of operator when applying pressure to cause expiration. (*After* C. K. Drinker (1938), *Carbon Monoxide Asphyxia*. Oxford University Press.)

In the prone-pressure method pressure is applied to the lowest ribs, so that the abdominal contents are pushed up into the thorax to drive out the air. Inspiration is a passive process and occurs because of the elastic recoil of the chest and abdomen. In an asphyxiated person however, there is a rapid loss of muscle tone and passive inspiration is not very effective. Experiments carried out recently on anæsthetized men given curare to make them apnœic have shown that Schäfer's method cannot always maintain a sufficient arterial oxygen saturation to prevent cyanosis.

The Arm-lift-back-pressure Method (Holger Nielsen).—This method has largely superseded the previous method and is now widely recommended. The patient is laid prone and foreign bodies are removed as quickly as possible from his mouth. If he has been submerged stand astride his thighs, clasp your hands under his abdomen and raise the patient up till water ceases to flow out of the mouth.

(1) Put the patient in the prone position shown in Fig. 32, 23A with the elbows flexed so that one hand rests on the other and one cheek rests on the uppermost hand.

(2) Kneel with one knee near the patient's head, placing the other foot near his elbow.

(3) Grip the patient's arms just above the elbows as shown in Fig. 32, 23A and rock backwards to the position shown in Fig. 32, 23B raising his arms until resistance is felt at the patient's shoulders. This expansive phase should take 2½ seconds.

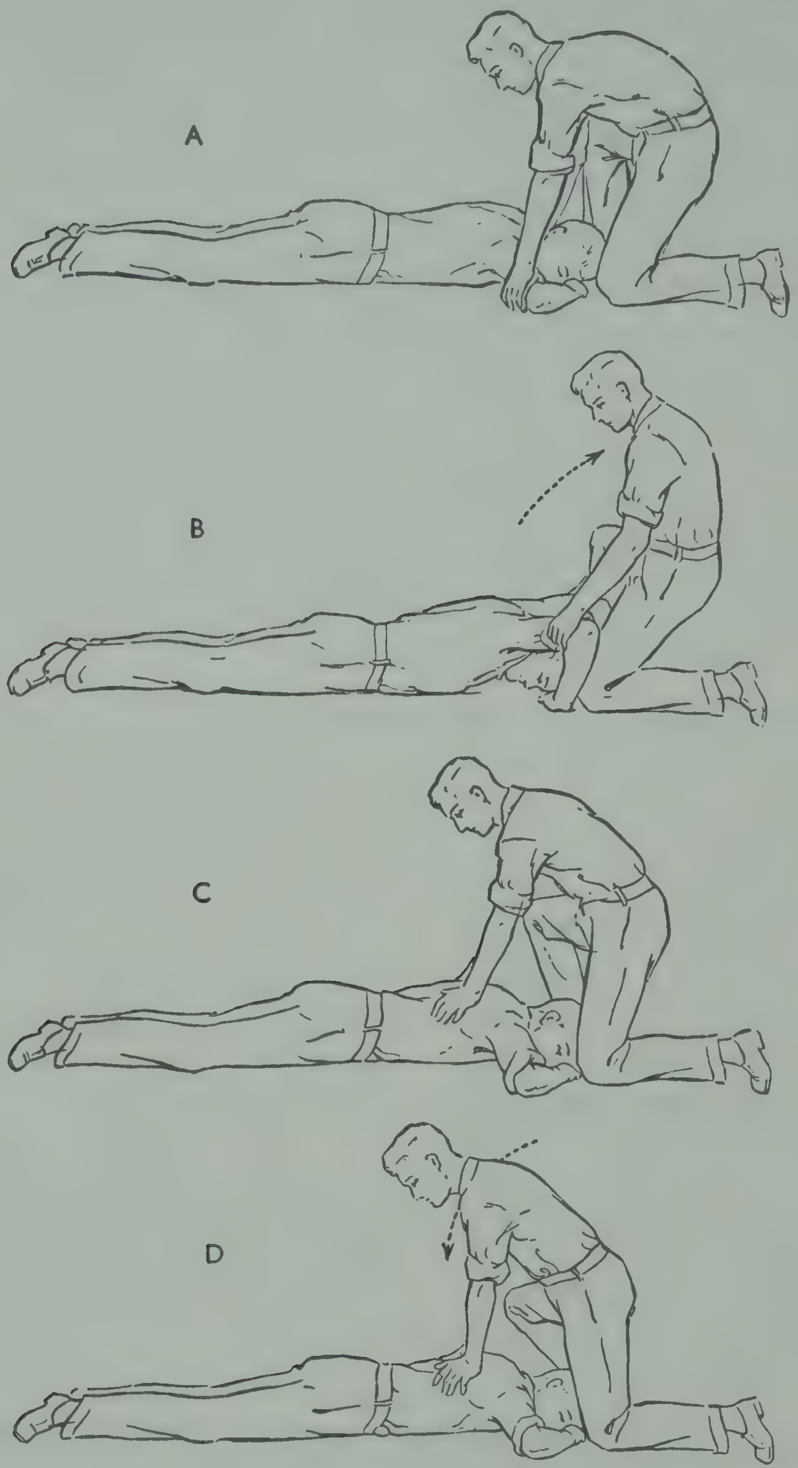


Fig. 32, 23

Holger Nielsen method of artificial respiration.

(4) Let the patient's arms drop then place your hands on his back just below the scapulæ (Fig. 32, 23C). Now rock forward, keeping your elbows straight, till your shoulders are nearly vertically over your hands as in Fig. 32, 23D. This compression phase should last 2½ seconds.

(5) Return without loss of time to the expansion phase described in (3) so that the complete cycle is repeated twelve times per minute. Keep up respiration by relays of helpers for at least two hours. If natural respiration begins synchronize your movements with the patient's movements and then gradually reduce the force of the artificial respiration.

Note that in this method both the compression and expansion phases are actively (not passively) brought about by the operator. Experiments on anæsthetized and curarized men showed that the tidal volume so produced may be two or three times the resting tidal volume and, even more important, the arterial oxygen saturation could always be kept at safe levels.

Rocking Method.—In Eve's method the patient is placed prone on some form of stretcher and tilted head up 45°, then head down 45°, about twelve times per minute (Fig. 32, 24). If the stretcher is balanced the effort required is small and the method may be kept up without fatigue for long periods. This method has been found useful on board ship for resuscitating the apparently drowned. Artificial respiration must be begun without delay by the arm-lift-back-pressure method and so maintained until the patient has been fixed in position on the rocking apparatus.

In the rocking method the air is moved in and out of the chest by the movement of the abdominal viscera. In the head-down position the weight of the viscera falls on the diaphragm which is pushed into the thorax and air is expelled. In the head-up position the abdominal viscera slide down again and air is drawn into the lungs. The rocking method appears to have a favourable effect on the circulation by assisting the venous return from the legs in the head-down position and from the head in the head-up position.

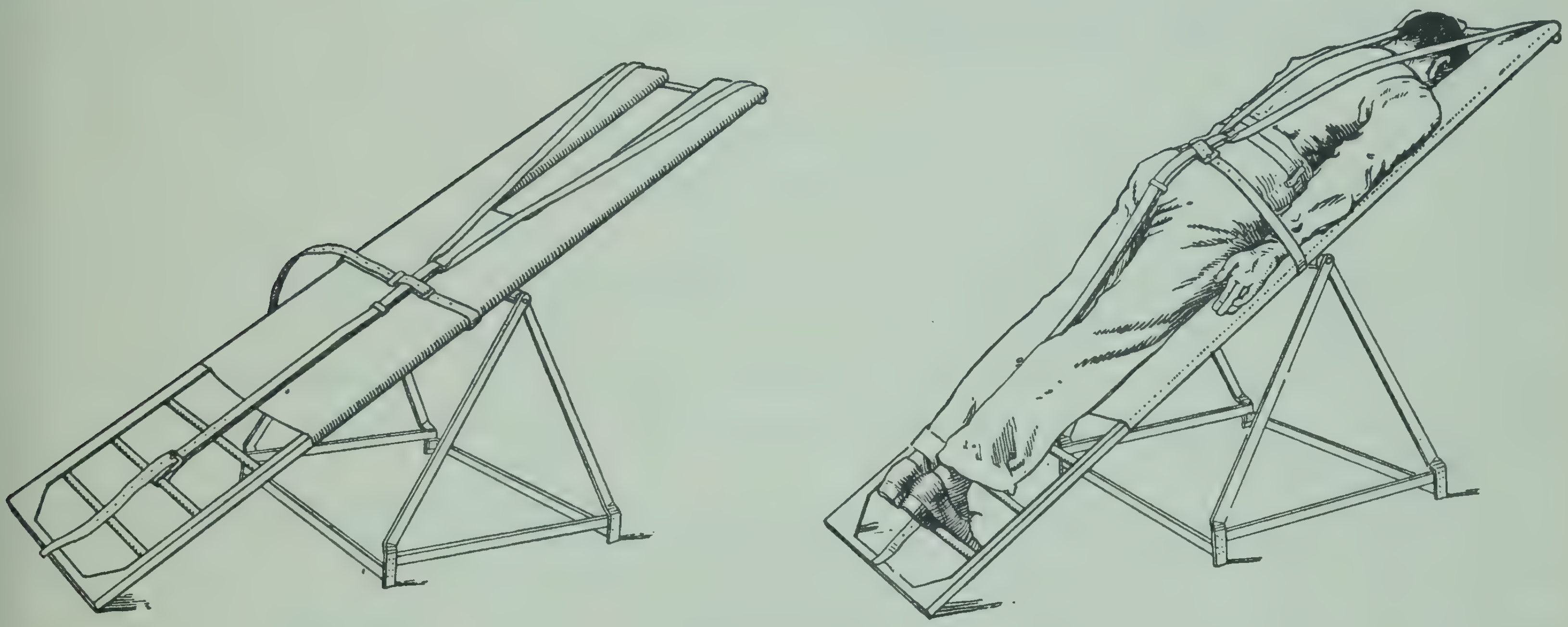


FIG. 32, 24

Rocking method of artificial respiration. (*After* E. M. Killick & F. C. Eve (1933), *Lancet*, ii, 741.)

In practising the methods it will be found that there is a tendency to apply pressure or to tilt too frequently. There is no advantage in this as the ventilation in the arm-lift-back-pressure and rocking methods is adequate, at the slow rate recommended. Opinion has changed recently about the wisdom of administering a mixture of oxygen and carbon dioxide in asphyxia. Inhalation of a mixture of oxygen and carbon dioxide to a healthy person increases the respiration so that there is practically no change in body carbon dioxide. If the respiratory centre is damaged or depressed, however, the appropriate increase in respiration may not take place and there may be a dangerous rise in the carbon dioxide in the body. High concentrations of carbon dioxide have an anæsthetic effect. Where the respiratory centre is still active the administration of additional carbon dioxide is unnecessary. For first-aid purposes oxygen only should be administered. Apart from the application of heat to maintain the body temperature and to allow hæmoglobin to part with its oxygen easily no other treatment is of real value.

Where prolonged artificial respiration is required, as sometimes in poliomyelitis, the patient is usually placed in a box with only his head protruding. A pump periodically (say ten times a minute) decreases the pressure in the box and so air is sucked into the patient's lungs. Between strokes of the pump the air is allowed into the box from the atmosphere, the chest recoils elastically and air is expelled from the lungs. Respiration can also be brought about by rhythmic electrical stimulation of the phrenic nerves through electrodes placed on the skin lying over them. Artificial respiration is nowadays often accomplished by an electrically driven pump which inflates the patient's lungs at a predetermined positive pressure through a tube inserted either through the mouth or nose or directly through a small incision into the trachea.

REFERENCES

Comroe, J. H., Forster, R. E., Dubois, A. B., Briscoe, W. A. J. & Carlsen, E. (1955). *The lung. Clinical physiology and pulmonary function tests.* Chicago: Year Book Publishers.

Donald, K. W. (1947). Oxygen poisoning in man. *Brit. med. J.* i, 667–672 and 712–717.

Donald, K. W. (1953). Definition and assessment of respiratory function, *Brit. med. J.* i, 415–422 and 473–478.

Donald, K. W. & Paton, W. M. D. (1955). Gases administered in artificial respiration with particular reference to the use of carbon dioxide. *Brit. med. J.* i, 313–318.

Dornhorst, A. C. (1955). Respiratory insufficiency, *Lancet,* i, 1185–1187.

Drinker, C. K. (1954). *Clinical physiology of the lungs.* Oxford: Blackwell.

Fowler, W. S. (1952). Intrapulmonary distribution of inspired gas. *Physiol. Rev.* **32,** 1–20.

Fulton, J. F. (1951). *Decompression sickness, Caisson sickness, Diver's and Flier's bends and related Syndromes.* London: Saunders.

Gilson, J. C. & Hugh-Jones, P. (1955). Lung function in coalworkers' pneumoconiosis. *Spec. Rep. Ser. med. Res. Coun., Lond.* No. 290. London: H.M.S.O.

Gray, J. S. (1950). *The Pulmonary Ventilation and its Physiological Regulation.* Springfield, Illinois: Thomas.

Haldane, J. S. & Priestley, J. G. (1935). *Respiration.* 2nd ed. Oxford: Clarendon Press.

Miller, W. S. (1947). *The Lung.* Springfield, Illinois: Thomas.

Pugh, G. & Ward, M. (1953). Appendix VII of *The Ascent of Everest* (1953), by John Hunt. London: Hodder & Stoughton.

Stickney, J. C. & Van Liere, E. J. (1953). Acclimatization to low oxygen tension. *Physiol. Rev.* **33,** 13–34.

Wallace, E. J. G. (1952). Resuscitation of the apparently drowned. *Brit. med. J.* ii, 278–280.

Whittingham, H. E. (1955). Medical science and problems of flying. *Brit. med. J.* i, 303–309.

CHAPTER 33

RESPIRATORY FUNCTION OF THE BLOOD

SINCE carbon dioxide appears in greater concentration in the expired air than in the inspired atmospheric air, it was at one time supposed that combustion took place in the lungs. This notion was shown to be false by H. G. Magnus about a hundred years ago, when he demonstrated that the blood leaving the lungs contained less carbon dioxide and more oxygen than the blood reaching them. In other words, the blood acts as a transport system for these two gases, bringing carbon dioxide from the tissues and carrying oxygen to the tissues. The fact that combustion takes place in the tissues can nowadays be shown easily by demonstrating the uptake of oxygen and output of carbon dioxide of tissue slices in a respirometer such as that devised by Warburg.

SOLUTION OF GASES IN WATER.—Before discussing the way in which the gases are carried in blood the conditions governing their solution in water must be mentioned. If the pressure of a gas above a volume of water is kept constant, the amount of the gas in solution diminishes as the temperature is increased. However, as the body temperature is remarkably constant, we need for physiological purposes to consider only the factors which modify the solubility at constant temperature. If the temperature is held constant, the amount of a gas dissolved in water depends directly on its pressure, or, if the gas under consideration is mixed with other gases, on its partial pressure. When equilibrium is established between the gas and the water as many molecules of the gas enter the water as leave it. The tendency of a gas to leave the liquid is called its *tension*—it is well demonstrated by the effervescence which follows the opening of a bottle of aerated (carbonated) water. The tension of a gas in water is measured by finding the pressure (or partial pressure) of the gas in the atmosphere with which it is in equilibrium.

The partial pressure of a gas in a mixture of gases is obtained by multiplying the pressure of the mixture by the percentage of the gas in it. For example, the partial pressure of oxygen in dry atmospheric air (21 per cent. oxygen) at 760 mm Hg is $\frac{21}{100} \times 760 = 160$ mm Hg. At 38° C 100 ml. of water dissolves 2·3 ml. of oxygen, 51 ml. of carbon dioxide and 1·2 ml. of nitrogen when the pressure of each gas is 760 mm Hg. Since arterial blood is nearly in equilibrium with alveolar air the tensions of the gases in arterial blood (p. 583) are approximately the same as their partial

	Oxygen ml.	Carbon Dioxide ml.	Nitrogen ml.
Water in equilibrium with alveolar air . . .	0·3	2·6	0·9

pressures in the alveolar air. The table on the preceding page shows the amounts of gases held in solution by 100 ml. of water at 38° C exposed to alveolar air.

Now when blood is exposed to a vacuum the whole of its contained gas is removed and the average volumes from 100 ml. of blood are as follows:

	Oxygen ml.	Carbon Dioxide ml.	Nitrogen ml.
Arterial blood	19	50	0·9
Venous blood	10·6	58	0·9

Thus the amount of oxygen in the blood is greater than can be accounted for by simple solution; most must be in some combination. Similarly, the greater part of the carbon dioxide must be present in combination. Nitrogen, on the other hand, is in simple solution.

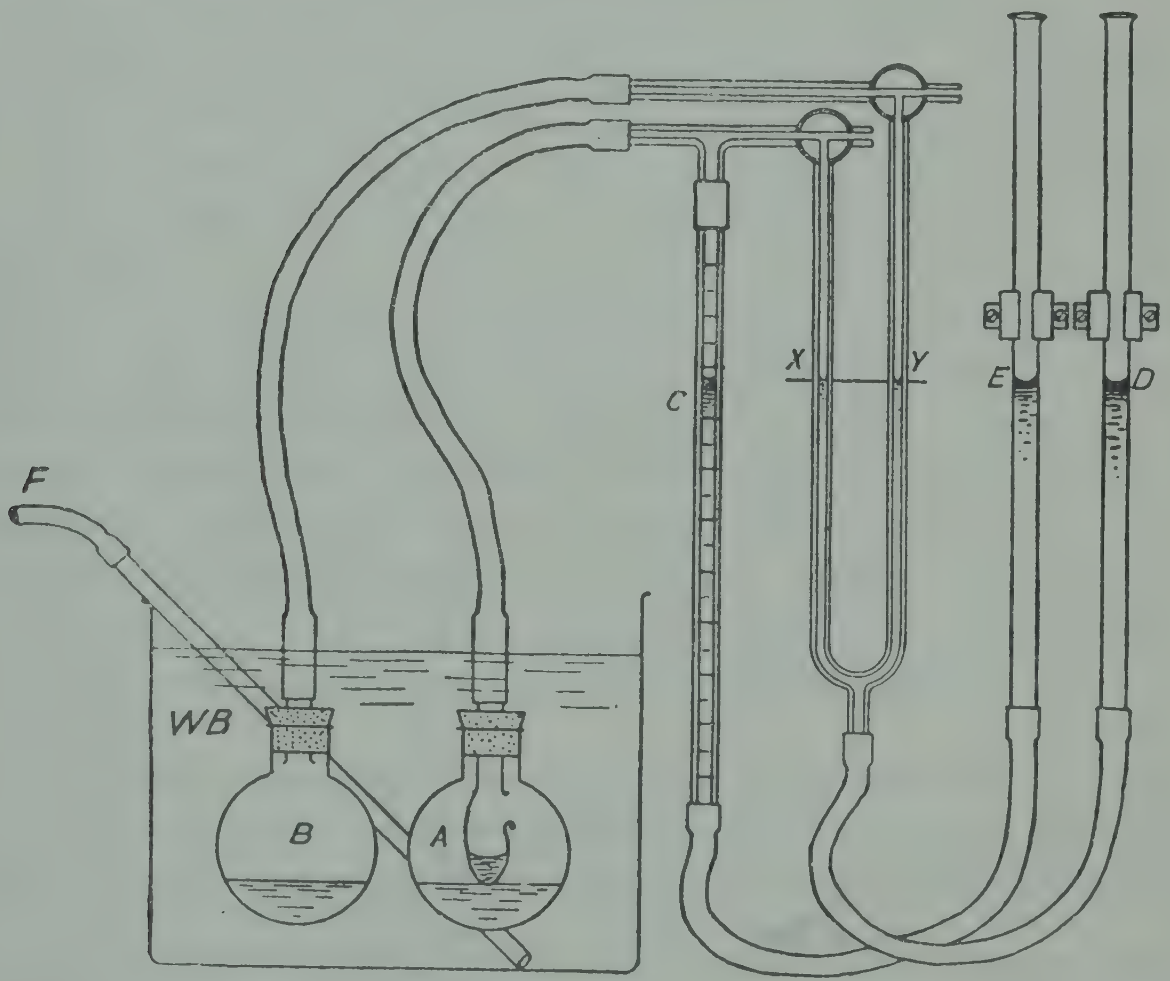


FIG. 33, 1

Haldane's blood gas apparatus. The reaction vessel A contains blood in a borate buffer. The inner vessel of A contains ferricyanide. The reaction vessel B, containing water, acts as a compensating thermobarometer. If vessel A is tilted the ferricyanide mixes with the blood, oxygen is evolved and depresses the level in the burette C. X, Y are gauge tubes, E, D levelling tubes. The water in the water-bath, WB, is stirred by air blown in through F (C. G. Douglas & J. G. Priestly (1949), *Human Physiology*, 3rd ed., p. 150. Oxford: Clarendon Press).

ESTIMATION OF OXYGEN AND CARBON DIOXIDE IN BLOOD

The oxygen and carbon dioxide content of blood can be estimated by Haldane's blood gas apparatus (Fig. 33, 1). It is necessary to introduce the oxalated blood sample into the apparatus without exposure to the atmospheric air in order to avoid changes due to loss of carbon dioxide or gain of oxygen, and the blood is therefore collected under mineral oil. After the red cells have been hæmolysed by saponin, potassium ferricyanide is added and all the oxygen is liberated (p. 408). This volume of oxygen actually present in the sample is its *oxygen content.* If the estimation is repeated on another sample of the same blood which has first been shaken with air to saturate it with oxygen, the figure obtained is higher than the oxygen content of the first sample and represents the *oxygen capacity.* The *percentage saturation* of the first sample is given by $\frac{\text{Oxygen content}}{\text{Oxygen capacity}} \times 100$. The carbon dioxide content of a blood sample can be measured in the same apparatus by adding tartaric acid, which drives off all the carbon dioxide.

The Van Slyke apparatus (Fig. 33, 2) is more frequently used for blood gas analysis. In this case the blood is put into an acidified air-free ferricyanide solution and subjected to a vacuum. The ferricyanide drives off the oxygen, the acid drives off the carbon dioxide and the vacuum the nitrogen. The gas evolved is analysed by absorbing first the carbon dioxide and then the oxygen.

Since the light transmissions, as shown in Fig. 25, 2, of hæmoglobin and oxyhæmoglobin are different, the oxygen saturation of blood can be approximately determined by shining a light through the lobe of the ear on to two photo-cells one of which is covered with a red filter and the other with a blue filter (the *oximeter*). The ratio of the photo-cell currents is related to the saturation of the blood with oxygen. This method, although approximate, is very rapid and convenient; it has been successfully used in the investigation of anoxia during flying at high altitudes and is likely to find increasing application in clinical medicine.

TRANSPORT OF OXYGEN

The amount of oxygen which a given amount of hæmoglobin takes up, that is its percentage saturation, is a function of the partial pressure of the oxygen pO_2 in the atmosphere in contact with it. The graph (Fig. 33, 3) showing the relation between the two, the *dissociation curve,* is obtained in the following way. A small quantity of blood, say 2 ml., is placed in a 250 ml. bottle (tonometer) which is filled with gas of known oxygen concentration. The bottle is turned on its side and rotated for twenty minutes in a water-bath at 37·5° C, so that a large area of blood is exposed to the gas. Then the percentage saturation is measured as described above. The results of a number of experiments at different pO_2 can be plotted to give Fig. 33, 3.

If a solution of hæmoglobin which has been dialysed against distilled water is used a rectangular hyperbola is obtained like curve A in Fig. 33, 3. A curve of this shape would be expected from the law of mass action if the reaction between oxygen and hæmoglobin were to be represented thus:

$$Hb + O_2 \rightleftharpoons HbO_2$$

It was discovered by Bohr in 1910 that if whole blood is used instead of a solution of hæmoglobin the curve is S-shaped, like curve B in Fig. 33, 3.

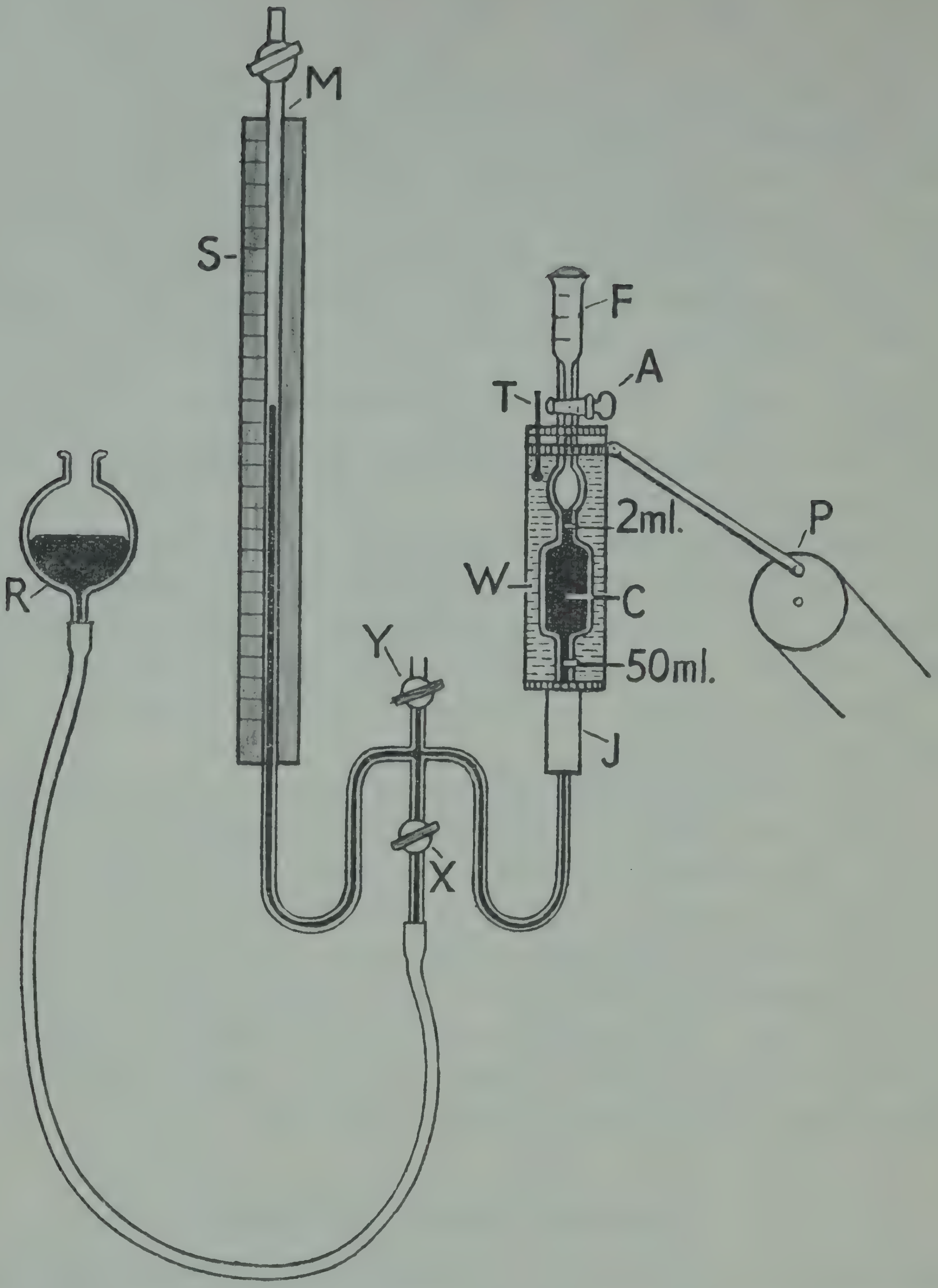


Fig. 33, 2

Van Slyke manometric apparatus. The apparatus consists essentially of a manometer tube, M, filled with mercury, with a scale, S, behind it. The level of mercury in the tube can be altered by raising and lowering the reservoir, R, when the tap, X, is open. The mixing chamber, C, is surrounded by a water-jacket, W, containing a thermometer, T, and can be shaken by means of a connecting rod attached to the pulley, P, which is driven by a small motor. The flexible joint, J, allows movement of the chamber and water-jacket while the rest of the apparatus remains steady. Blood or other fluid may be introduced into the mixing chamber through the funnel, F, and the tap, A. Below the lower bulb of the mixing chamber, C, is a graduation mark at 50 ml., and between the large and the small bulb is a graduation mark at 2 ml. When a gas is released from the fluid in C it can be compressed to a constant volume by raising the mercury level in the reservoir, R, until the mercury in C reaches the 2 ml. graduation mark. The pressure of gas in C can then be read off on the manometer tube, M.

Barcroft later showed that this change of shape was due to the electrolyte content of the blood. Bohr also found that the position of the dissociation curve depended on the partial pressure of carbon dioxide in the mixture of oxygen and carbon dioxide to which the blood sample is exposed in the tonometer—the greater the pCO_2 the further to the right was the dissociation curve (Fig. 33, 4). Barcroft reinvestigated the effect of carbon dioxide and found that the displacement of the curve to the right could be produced by any acid and was, in fact, dependent on pH. A rise of temperature has much the same effect as a rise in hydrogen-ion concentration, in other words the blood parts with its oxygen more readily at higher temperatures.

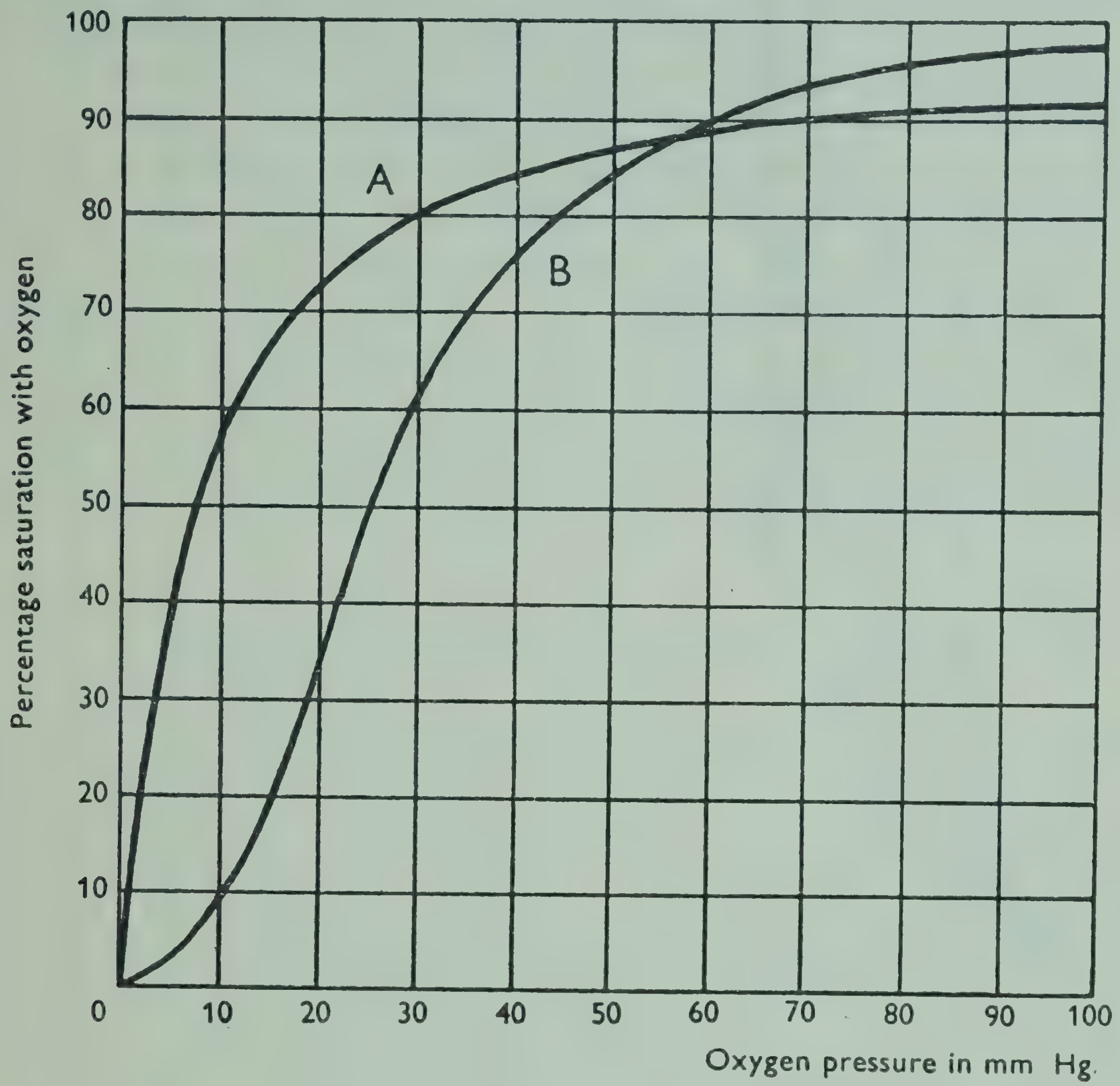


FIG. 33, 3

A, dissociation curve of oxyhæmoglobin in solution in water. B, dissociation curve of oxyhæmoglobin in blood at 40 mm Hg of pressure of carbon dioxide. (Redrawn from J. Barcroft (1914), *Respiratory Function of the Blood*, p. 160. Cambridge University Press.)

Physicochemical explanations of the shape of the dissociation curve of hæmoglobin as it exists in blood have been sought but a final conclusion has not yet been reached. The curve would be **S**-shaped if the hæmoglobin existed as an aggregate of several molecules of hæmoglobin, Hb_2, Hb_3 or Hb_4, each molecule containing one atom of iron. The molecular weight of hæmoglobin obtained by the ultracentrifuge or derived from measurement of the osmotic pressure is about 68,000, and therefore each molecule must have four iron atoms (p. 404). The law of mass action applied to such an aggregate, however, gives a much more accentuated **S**-shaped dissociation curve than is actually obtained.

(Fig.

The modification of the dissociation curve produced by the carbon dioxide tension of the blood (the Bohr effect) has biological advantages. In the lungs, where the oxygen tension is high and the carbon dioxide tension is low, the blood takes up more oxygen than it could were the carbon dioxide content high. When the blood passes to the tissues, where the oxygen tension is low and the carbon dioxide tension high, the hæmoglobin parts more easily with its oxygen. The more active the tissue the more carbon dioxide does it produce, and the increased carbon dioxide tension, since the dissociation curve is lowered and moved to the right, sets free more oxygen for the respiratory processes of the tissue. Thus the Bohr effect allows the blood to take up more oxygen in the lungs and facilitates the liberation of oxygen from the blood in the tissues.

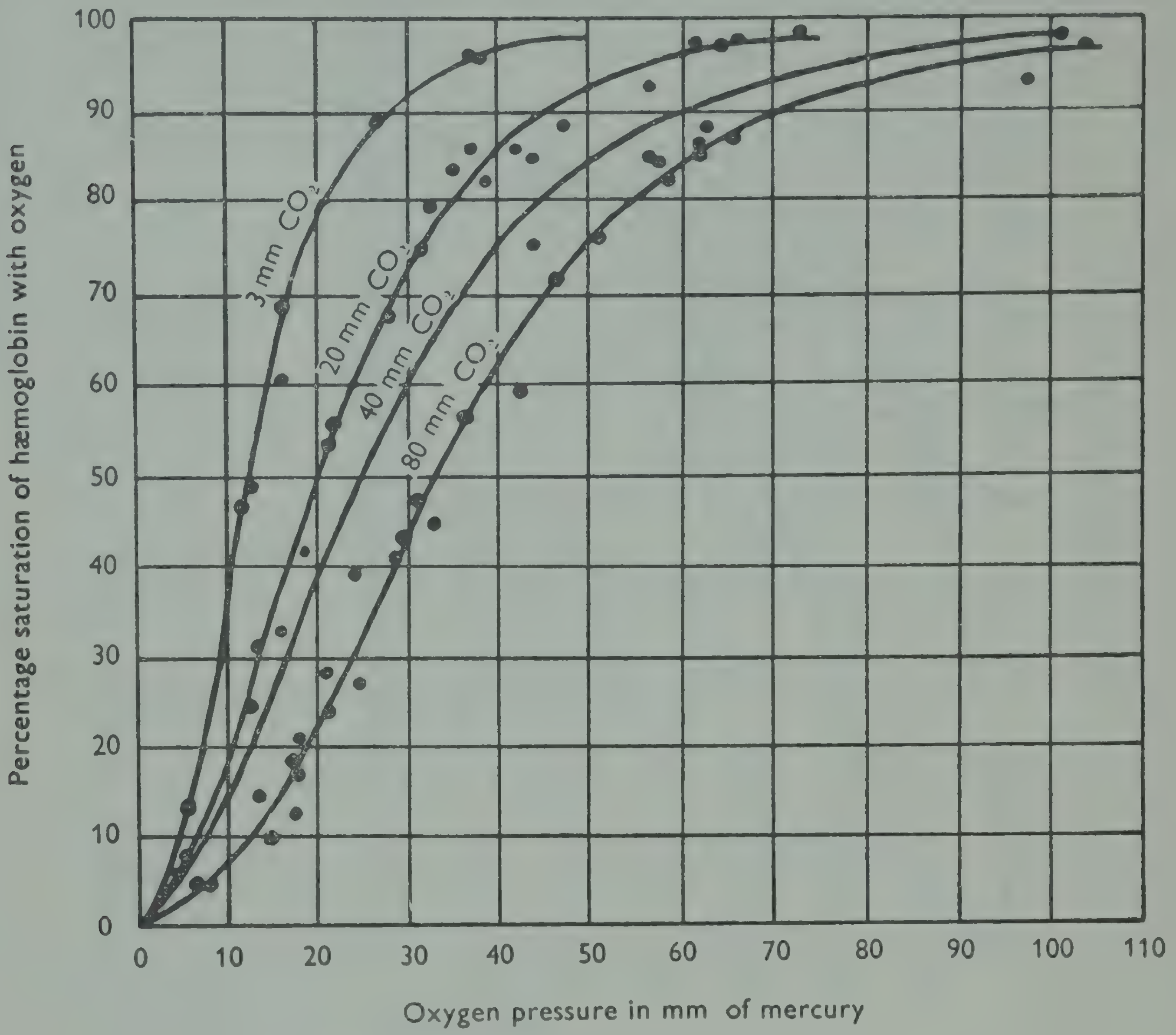


FIG. 33, 4

Oxygen dissociation curves of human blood exposed to pressures of 3, 20, 40, and 80 mm Hg of carbon dioxide. The percentage saturation was measured by Van Slyke's apparatus (A. V. Bock, H. F. Field & G. S. Adair (1924), *J. biol. Chem.* **59,** 366).

The arterial blood leaving the lungs is nearly in equilibrium with the alveolar air in which the tension of oxygen is about 105 mm Hg and that of carbon dioxide 40 mm Hg. By experiment in man it is found that the arterial oxygen tension is about 95 mm Hg and the carbon dioxide tension 40 mm Hg. Reference to the curve of Fig. 33, 4, shows that hæmoglobin is 95 per cent. saturated in these circumstances. In mixed venous blood the oxygen and carbon dioxide tensions are 40 and 46 mm Hg respectively. The hæmoglobin in mixed venous blood is, therefore, about 70 per cent. saturated.

The speed of the reaction $Hb + O_2 \rightarrow HbO_2$ has been measured by driving a hæmoglobin solution through a jet into a mixing chamber and water containing oxygen in solution through another jet into the same mixing chamber, from which the mixed fluids travel steadily down a glass observation tube. While the mixed fluids pass down the tube the hæmoglobin becomes oxyhæmoglobin. The point at which 50 per cent. saturation occurs can be detected spectroscopically (see p. 408). The rate of flow can be calculated from the diameter of the tube and the volume flowing out of it. The distance from the mixing chamber to the point of half-saturation divided by the rate of flow gives the time to reach half-saturation. This is between 0·001 and 0·01 seconds. This time is so much less than the half second or so spent by a corpuscle in the alveolar capillary that the speed of this reaction cannot be a factor limiting the uptake of oxygen. The rate of diffusion across the alveolar membrane may, however, impose a limit to the oxygenation of the blood (p. 584). The speed of the reverse reaction, $HbO_2 \rightarrow Hb + O_2$, although not quite so fast, is of the same order, so that release of oxygen to the tissues from the blood can be very rapid. The speed of unloading of oxygen is increased by a rise of temperature and fall of pH and these conditions are found in an exercising muscle.

THE CARRIAGE OF CARBON DIOXIDE

As indicated already, the carbon dioxide of the blood cannot be in simple physical solution. At 38° C and 40 mm Hg 100 ml. of water can dissolve about 2·6 ml. of carbon dioxide, whereas blood may contain 50 or 60 ml. carbon dioxide per 100 ml. The fact that all the carbon dioxide can be obtained from blood by addition of an acid suggests that it is present as a bicarbonate. Furthermore the total anions (chloride, phosphate, and others) are insufficient to combine with the sodium ions found in plasma. It is, therefore, most likely that the combined carbon dioxide is present as sodium bicarbonate. Both corpuscles and plasma transport carbon dioxide; the former carrying about one-third of the total and the latter about two-thirds.

In much the same way as with oxygen we can obtain data to draw a dissociation curve showing the relationship between the pressure of carbon dioxide to which a solution is exposed and the total amount of carbon dioxide contained in it. Fig. 33, 5, shows dissociation curves for a number of solutions. *Separated plasma* is obtained by centrifuging blood and pipetting off the supernatant fluid; the plasma is then exposed to various pressures of carbon dioxide and its total carbon dioxide content measured. The amount of carbon dioxide taken up (as dissolved carbon dioxide and as bicarbonate) obviously depends on the properties of the plasma alone. If whole blood is first exposed to a known pressure of carbon dioxide and is then centrifuged under paraffin to prevent loss or gain of gas the plasma can then be removed and its carbon dioxide estimated. The carbon dioxide in this *true plasma* is obviously dependent partly on reactions in the plasma and partly on those occurring in the corpuscles.

The uppermost curve in Fig. 33, 5, shows that when a solution of sodium arbonate is exposed to a vacuum (zero pCO_2) half the total carbon dioxide content is retained, the other half being removed according to the equation:

$$2NaHCO_3 \rightarrow Na_2CO_3 + CO_2 + H_2O$$

If, however, the same solution is treated with an acid all the carbon dioxide is liberated thus:

$$2NaHCO_3 + 2HCl \rightarrow 2NaCl + 2CO_2 + 2H_2O$$

Separated plasma behaves like a bicarbonate solution except that rather more than half of the total carbon dioxide is brought off by a vacuum.

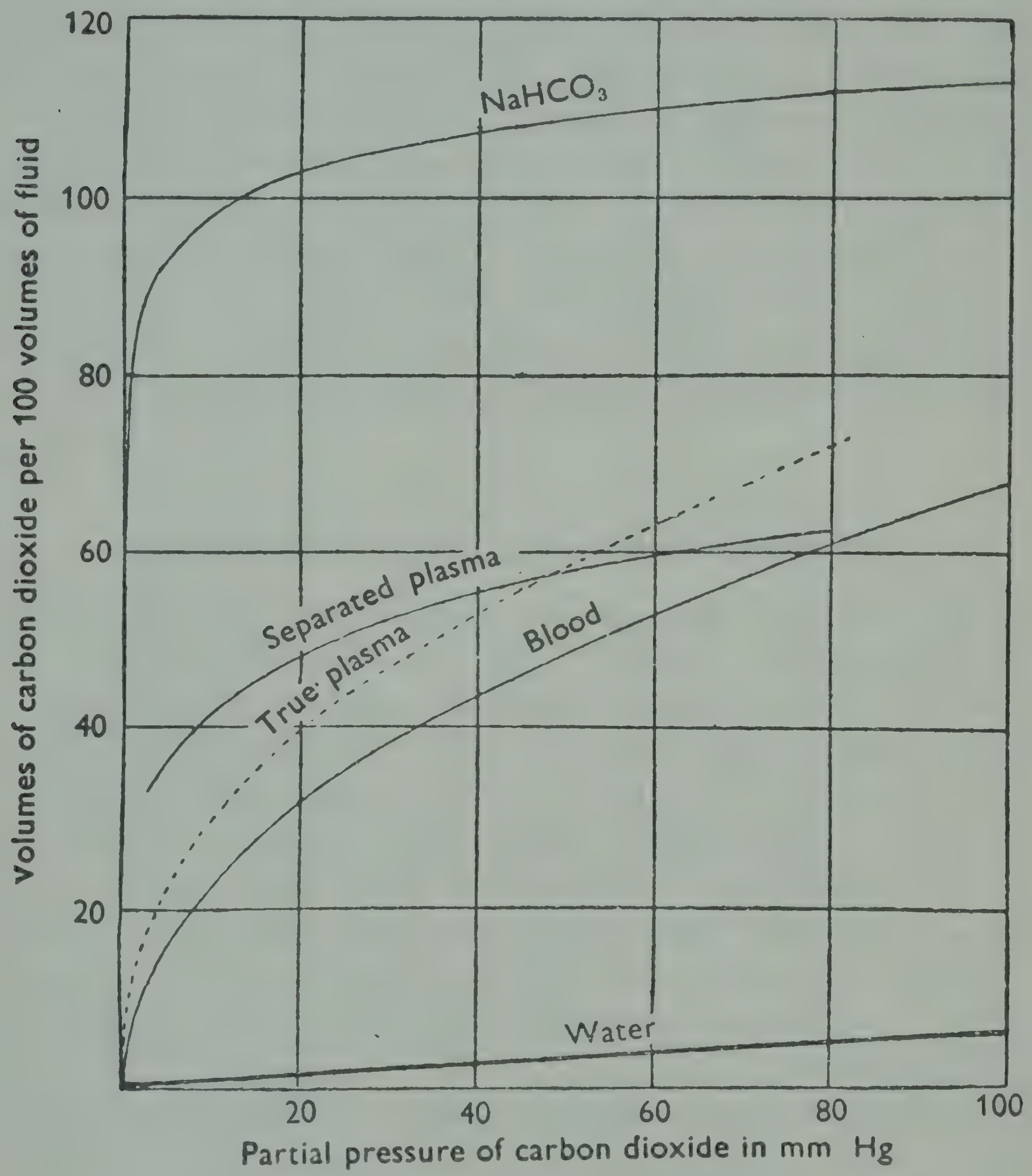


FIG. 33, 5

Carbon dioxide dissociation curves for water, blood, separated plasma, true plasma and 0·0484 N $NaHCO_3$ which contains the same amount of available base as does blood. (*After* G. M. Wishart and C. Lovatt Evans from data in T. R. Parsons (1919), *J. Physiol.* **53,** 42, 340, and J. Joffe & E. P. Poulton (1920), *J. Physiol.* **54,** 133.)

The curves for true plasma and for blood show that at zero pCO_2 no carbon dioxide is left. Obviously there must be some process, due mainly to the corpuscles, which aids the liberation of carbon dioxide. This can be confirmed by adding red cells to separated plasma after it has been exposed to a vacuum; a further amount of carbon dioxide is evolved, as if the corpuscles contained an acid.

Chloride Shift.—The main chemically measurable changes occurring in blood on exposure to carbon dioxide were described by Hamburger ma years ago. In both plasma and corpuscles the bicarbonate content increases; the chloride content of the plasma diminishes but in the red cells the chloride content increases. This *chloride shift*, or Hamburger phenomenon, is actually a movement of ions, but it may be easier to understand it from the following equations. When carbon dioxide reaches the blood the following reaction takes place:

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

The blood tends to become more acid, but the change is very small because of the buffering action of the proteins and phosphate in the plasma and of the hæmoglobin and phosphate in the corpuscles. Since the blood at pH 7·4 is on the alkaline side of the iso-electric points of both plasma proteins and hæmoglobin, these proteins are acting as weak acids. The hæmoglobin within the corpuscle is attached chiefly to potassium and the proteins of the plasma to sodium. Since the KHb and Na protein are salts of weak acids the cation (K^+ or Na^+) is easily displaced by stronger acids. In the plasma the following reaction occurs,

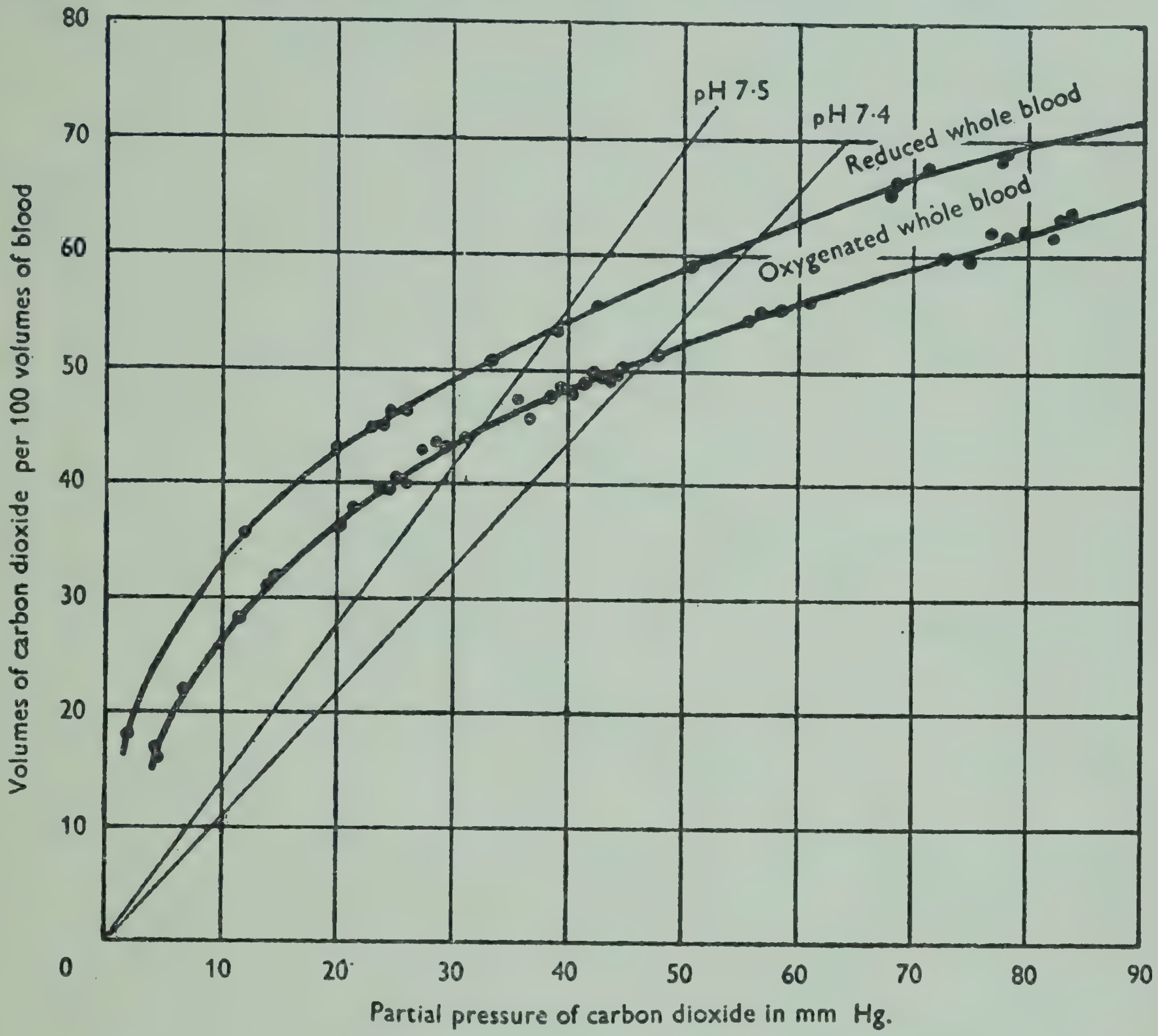


FIG. 33, 6

Carbon dioxide dissociation curves for reduced and oxygenated blood. (*After* A. V. Bock, H. F. Field & G. S. Adair (1924), *J. biol. Chem.* **59**, 371.)

$$H_2CO_3 + \text{Na proteinate} \rightarrow NaHCO_3 + \text{H proteinate}$$

In the red cell, since the corpuscular envelope is freely permeable to carbon dioxide,

$$CO_2 + H_2O \rightarrow H_2CO_3 \rightarrow H^+ + HCO_3^-$$

$$H_2CO_3 + KHb \rightarrow KHCO_3 + HHb.$$

(Fig.

Since the buffering action of the hæmoglobin is greater than that of the plasma, HCO_3^- ions are formed in the corpuscles in greater amount than in the plasma. Because of this concentration gradient HCO_3^- ions pass out of the red cells into the plasma and Cl^- ions enter the cells.

Oxyhæmoglobin is a more acid substance than hæmoglobin, that is it gives up its base less readily and therefore the reaction

$$H_2CO_3 + KHbO_2 \rightarrow KHCO_3 + HHbO_2$$

does not occur so easily as the corresponding one with KHb. The result is, as shown in Fig. 33, 6, that the more oxygen the blood holds the less carbon dioxide it can hold and vice versa. When the venous blood is oxygenated in the lungs the processes involved in the chloride shift just described are reversed.

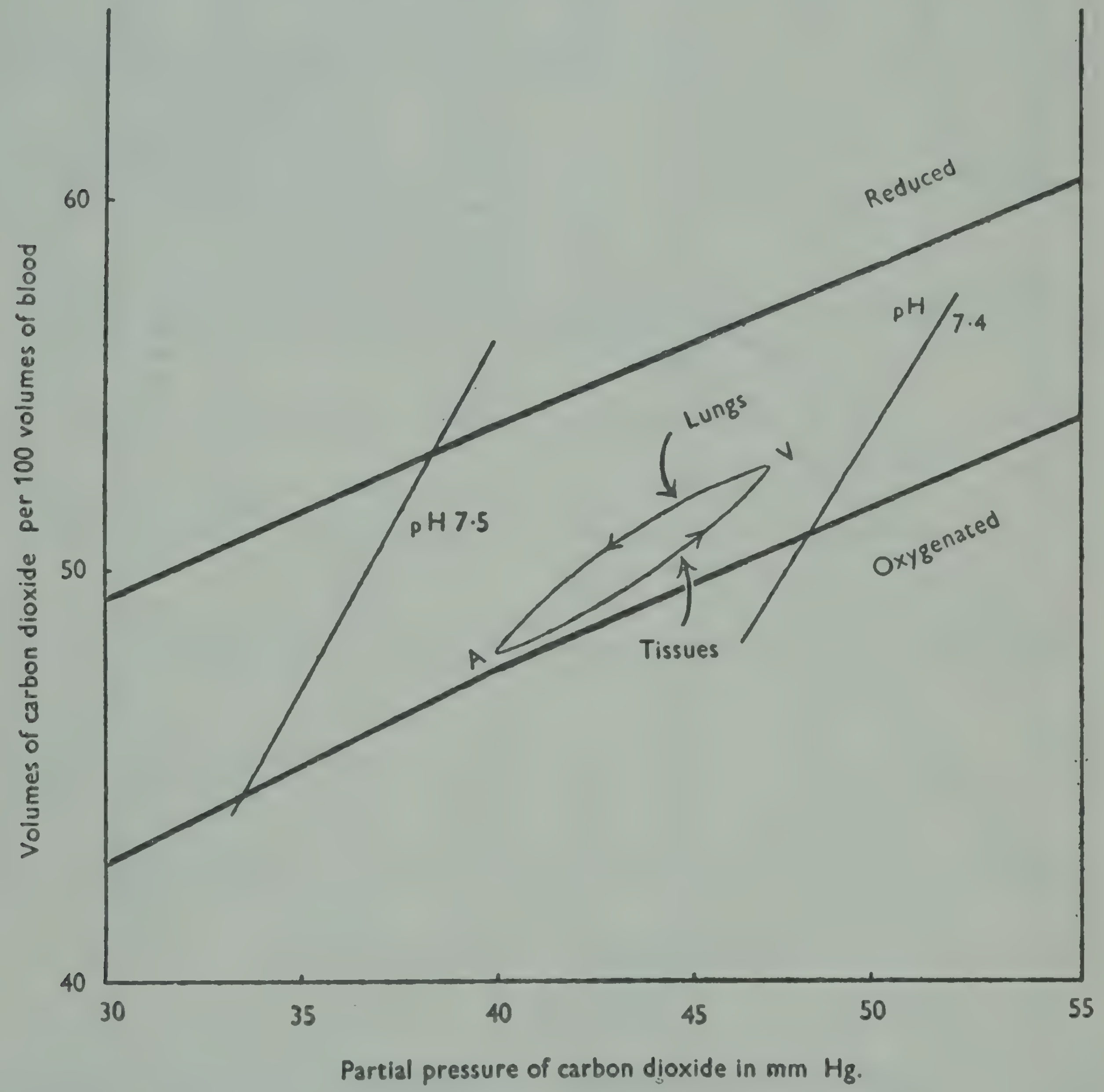


FIG. 33, 7

An estimate of the dynamic changes from arterial to venous blood. The loop AV allows for the diffusion process in the lungs and tissues. Arterial blood 40 mm Hg carbon dioxide, 95·5 per cent. saturated. Mixed venous blood 47 mm Hg carbon dioxide, 65 per cent. saturated. (*After* L. J. Henderson, A. V. Bock, H. F. Field & J. L. Stoddart (1924), *J. biol. Chem.* **59**, 379.)

The portion of the diagram which is of physiological interest is given on an enlarged scale in Fig. 33, 7. Since the blood is never completely reduced, nor fully oxygenated, the changes occurring in the lungs or tissues must be somewhere between the two lines; in the subject whose blood provided th data for Fig. 33, 6, and Fig. 33, 7, the loop AVA gives an approximation to the changes in the lungs and tissues allowing for the diffusion of the gases through them.

In the tissues the result of all these changes is an increase in the osmotic pressure of the red cells which is reduced by intake of water molecules. The increase in size of the red cells makes the venous hæmatocrit greater than the arterial.

Carbonic Anhydrase.—Although in pure solution the reaction

$$CO_2 + H_2O \rightarrow H_2CO_3$$

is very slow, the subsequent reactions mentioned above, being ionic, are very rapid. Roughton discovered that the red cells possess an enzyme which he called *carbonic anhydrase* (see also p. 240). This enzyme increases greatly the speed of the reaction $CO_2 + H_2O \rightleftharpoons H_2CO_3$ in either direction, so that the short time of transit of the blood through the lungs is sufficient for the release of its carbon dioxide. Conversely the uptake of carbon dioxide from the tissues is greatly facilitated. The enzyme is found in the corpuscles but not in the plasma. Thus when carbon dioxide leaves the blood plasma and passes into the alveolar air it is replaced chiefly from the carbon dioxide of the corpuscles, because the possession of carbonic anhydrase by them means that the bicarbonate of the corpuscles can yield carbon dioxide more quickly than the bicarbonate of the plasma. When blood flows through the tissues it is exposed to a high pressure of carbon dioxide; this gas diffuses into the plasma and then into the red cells, where carbonic acid is formed much more rapidly than in the plasma. Carbonic acid quickly ionizes and the concentration of bicarbonate ions increases, they leave the cell and chloride ions enter as explained on the previous page.

Carbamino Compounds. — The action of carbonic anhydrase can be inhibited by cyanide and sulphonamides. Blood with a trace of cyanide added still takes up a small amount of carbon dioxide at a rate which is greater than that shown by pure bicarbonate solutions. This is accounted for by the existence of carbamino compounds formed thus:

$$CO_2 + PrNH_2 \rightleftharpoons PrNH.COOH$$

where $PrNH_2$ is a protein with a free amino group. In this combination the hæmoglobin carries between 5 and 10 per cent. of the total carbon dioxide of the blood (Fig. 33, 8). This carbamino-bound carbon dioxide although small in amount is very mobile and is, therefore, of considerable physiological importance. As shown in Fig. 33, 8, about one-third of the change in carbon dioxide content of the blood in ordinary respiration can be accounted for by the change in the carbamino-bound carbon dioxide. The measurements of total carbon dioxide content quoted above (p. 577) included this carbamino fraction, which is affected by the oxygen tension in the same way as the carbon dioxide carried as bicarbonate; that is to say, reduced hæmoglobin can carry more carbamino-bound carbon dioxide than oxyhæmoglobin.

The factors involved in the transport of carbon dioxide by the red blood cell have been summarized by Roughton in the form of a diagram (Fig. 33, 10).

Passage of the Gases through the Alveolar Membrane.—There was at one time considerable discussion as to whether the passage of oxygen from alveolar air to blood, and of carbon dioxide in the opposite direction, could be completely accounted for by diffusion, or whether some secretory process was involved. To calculate if diffusion can account for the passage of the gases it is necessary to know (1) the partial pressures of the gases in the alveolar air, in the blood of the pulmonary artery and pulmonary veins, and (2) the rate of diffusion of gases through the alveolar membrane and the distances involved.

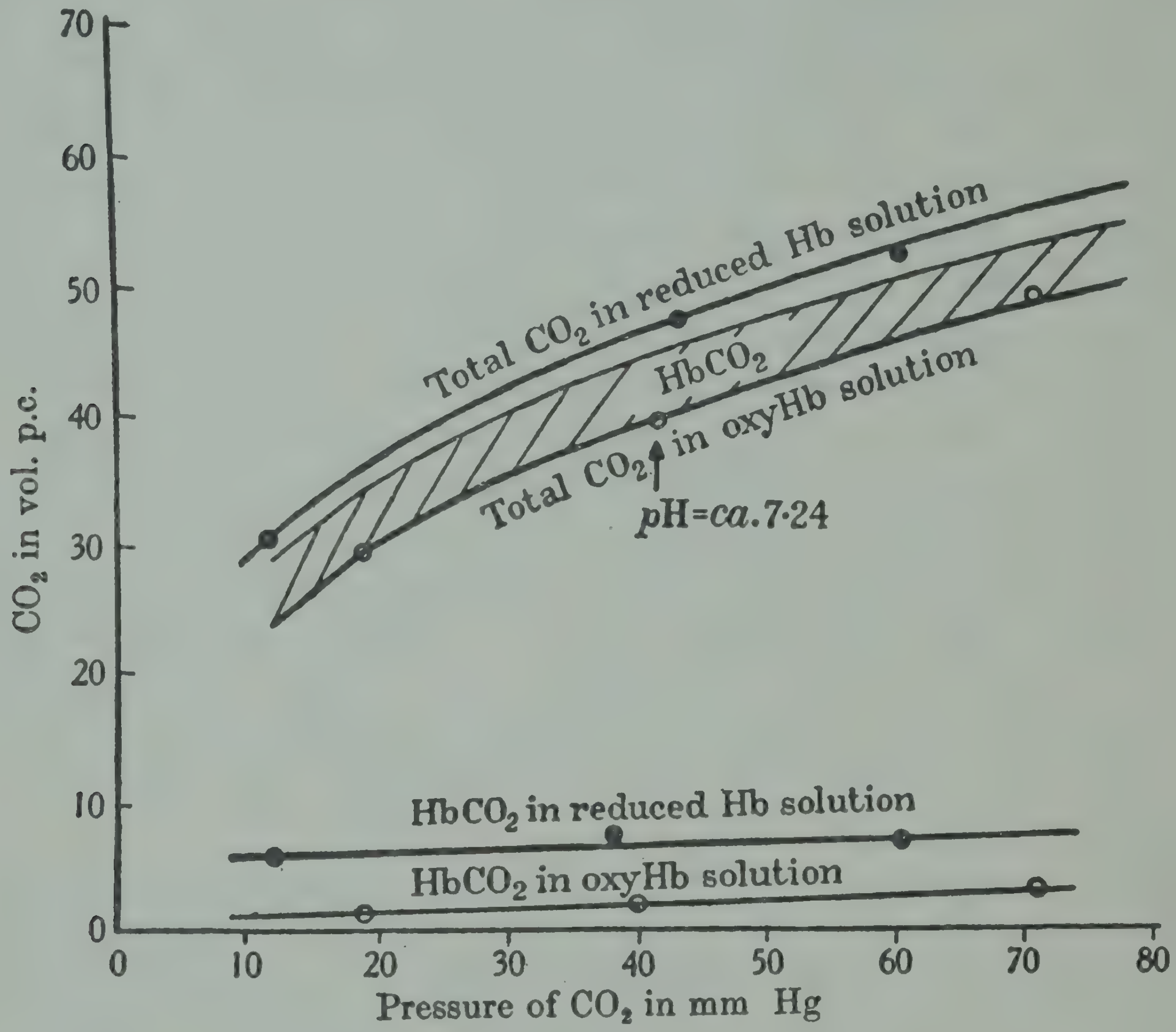


Fig. 33, 8

Upper two curves are carbon dioxide dissociation curves of reduced and oxygenated human hæmoglobin solution respectively. In the change from the oxygenated to the reduced state the carbon dioxide represented by the shaded area is taken up in the carbamino form—this is quite a considerable part of the change in the carbon dioxide content. The absolute amounts of carbamino compound are shown in the lower two curves. (J. K. W. Ferguson (1937), *J. Physiol.* **88,** 49.)

The methods for obtaining samples of alveolar air and for calculating the partial pressure of the gases are given on pp. 551 and 571. In experiments on animals the tensions of the gases in the blood are measured by a tonometer (Fig. 33, 11) which is connected between the two ends of a divided artery or vein in such a way that the blood flows past a bubble of air trapped in the upper part of the apparatus. When a large amount of blood has gone through the apparatus equilibrium between the gases in the blood and in the air bubble is reached. From the analysis of the air bubble the partial pressures of the gases in the bubble and the tension in the blood can be calculated.

Krogh showed that the carbon dioxide tension of the arterial blood of the rabbit was nearly the same as the partial pressure of carbon dioxide in the alveolar air; usually the former was slightly the greater. Carbon dioxide diffuses so rapidly (about thirty times more rapidly than oxygen) that a pressure difference of only 0·03 mm Hg is necessary to account for the passage across the alveolar membrane of carbon dioxide in the amounts required (p. 544), and accordingly in the case of carbon dioxide a process of diffusion is adequate to account for the facts of respiration.

Recent measurements show that the tension of oxygen in the peripheral arterial blood in man at rest is on the average 9 mm Hg below that in the alveolar air. During severe exercise it has been found that this pressure difference increases up to 16·3 mm Hg. The difference in oxygen tension between the alveolar air and the blood leaving the alveolar capillaries is in fact much less than this, because after the oxygenated blood has left the alveoli it becomes mixed with a small amount, estimated to be about 4 per cent. of the cardiac output, of venous blood which comes from the bronchial vessels (p. 539), perhaps from poorly ventilated alveoli and also from a few coronary veins which open into the left side of the heart (p. 426). It has been calculated that the difference in partial pressure of oxygen across the pulmonary membrane (the membrane gradient) is of the order of 1 mm Hg. The venous blood has an oxygen tension of about 40 mm Hg and when it enters the alveolar capillary it is exposed to alveolar air with an oxygen tension of 105 mm Hg: as the blood flows through the capillary its oxygen tension rises and approaches that in the alveolar air up to a final value. This final value is the tension in the alveolar air less the membrane gradient of 1 mm Hg, that is, about 104 mm Hg. The membrane gradient is little altered in exercise, but is greatly increased at low barometric pressures at which the oxygen content of the blood is reduced. The value of the membrane gradient at 5000 metres (16,500 feet) at rest is about 8 mm Hg. This pressure head is needed at low barometric pressures to transfer the required volume of oxygen across the alveolar membrane.

Table 33, 9

Table summarizing the Differences between Arterial and Mixed Venous Blood

	Arterial Blood	Mixed Venous Blood
Oxygen		
— content (ml. per 100 ml.)	19	10·6
— tension (mm Hg)	95	40
— saturation (percentage)	95	70
Carbon dioxide		
— content (ml. per 100 ml.)	50	58
— tension (mm Hg)	40	46

In exercise the oxygen requirements are greatly increased and large quantities of oxygen must pass across the alveolar membrane. There is in fact a considerable enlargement of the pulmonary capillary bed but, although the overall rate of blood flow through the pulmonary circuit is much increased, the time spent by each red corpuscle in passing through a pulmonary capillary is not so greatly reduced (p. 528). The oxygen saturation of arterial blood in exercise is as high as it is in the resting state unless the subject is in the last stages of exhaustion. The increased demand of the muscles is met partly by the opening up of capillaries closed in the resting state, by the greatly augmented blood flow through the muscle as the result of arteriolar dilatation and by an increase in the arterio-venous oxygen difference. In these circumstances the oxygen tension in the venous blood returning to the lungs is lower than normal. Owing to the greater depth of respiration the percentage of oxygen in the alveolar air may rise from the resting value of 14 to 17 or 18 so that the blood in the lungs is exposed to alveolar air of greater than normal oxygen tension. The pressure gradient responsible for diffusion of oxygen into the pulmonary blood is thus increased.

It is now, as the result of numerous experiments, generally agreed that the alveolar membrane acts in a purely passive manner and that secretion plays no part in the transference of oxygen or carbon dioxide. The characteristics of the alveolar membrane can be stated in terms of its *diffusing capacity* which is the number of ml. of a gas which can traverse it per minute per mm Hg gradient of pressure across the membrane. The value at rest is about 21 for oxygen and 17 for carbon monoxide (the latter is easier to measure). The speed with which oxygen can pass from alveolar air to the hæmoglobin of the red cells is obviously affected by the permeability of the structures crossed, the area over which the hæmoglobin is spread and the rate of uptake of oxygen by hæmoglobin. The diffusing capacity is increased in healthy persons in exercise up to three times. The resting value is reduced if the lung parenchyma is reduced as in emphysema and it may not increase on exercise.

Ventilation-Perfusion Relationship

It has long been known that when the lungs are diseased the alveolar air may vary in composition in different parts of the lungs and that an area of lung de-aerated by disease is often perfused with blood at a slower rate than the surrounding healthy lung. It is possible, to calculate the composition of the 'ideal' alveolar air, that is the composition of the alveolar air assumed to exist throughout the lung to account for the total gas exchange of the subject on the assumption that all the alveoli received an equal share of the inspired air and an equal share of the blood coming from the pulmonary artery. Since the quantity of oxygen taken up by the blood is the same as the quantity lost by the alveolar air, and the quantity of carbon dioxide added to the alveolar air is the same as that lost by the body, the R.Q. of the blood and alveolar air are the same. If part of the lung is poorly ventilated but well perfused with blood the blood leaving this part will be poorly oxygenated and its composition will be nearly the same as mixed venous blood and the R.Q. of the alveolar air sample will be low. If part of the lung is well ventilated but poorly

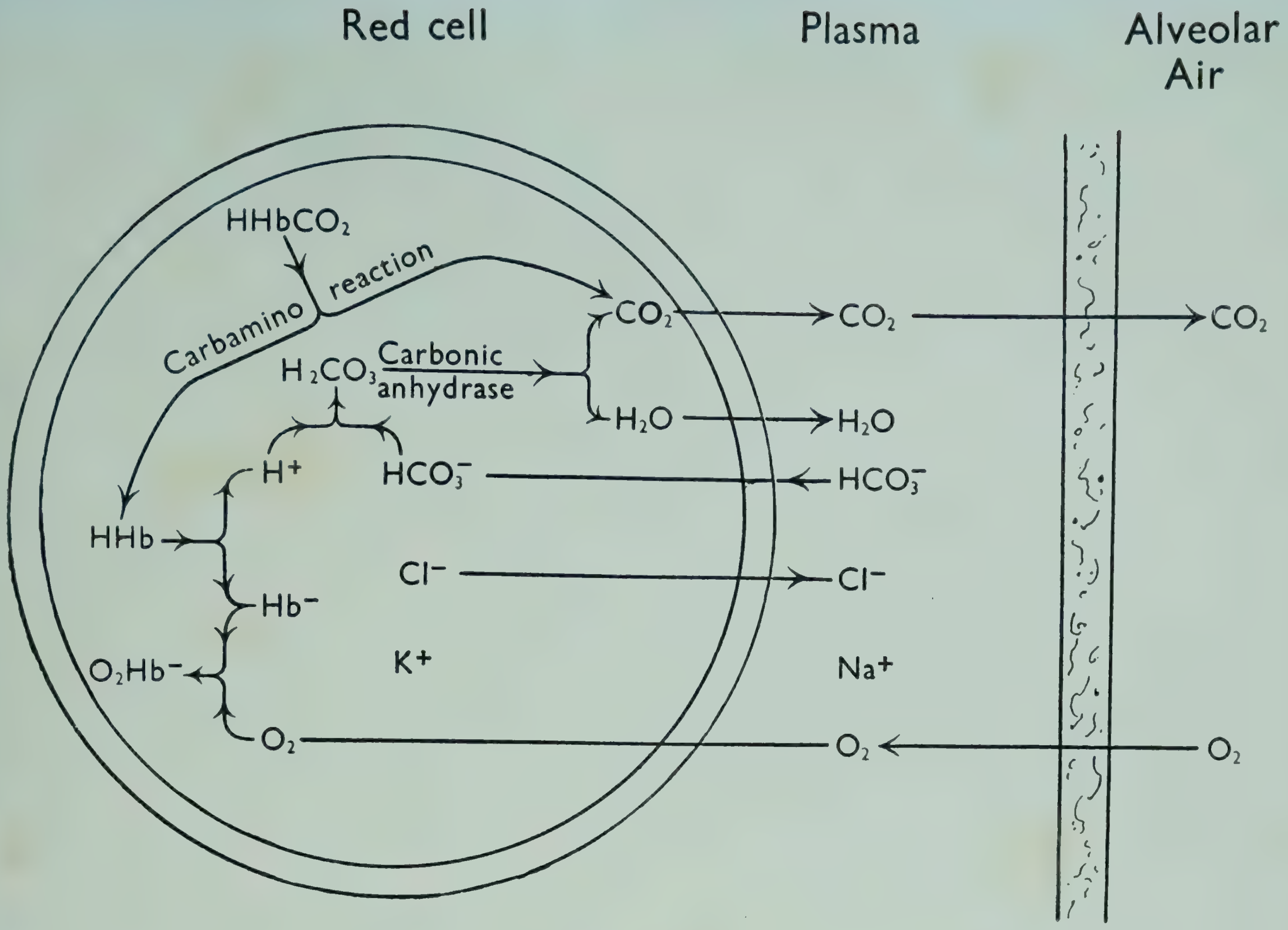

Red cell
Plasma
Alveolar Air
HHbCO₂
Carbamino reaction
CO₂
CO₂
CO₂
H₂CO₃
Carbonic anhydrase
H₂O
H₂O
H⁺
HCO₃⁻
HCO₃⁻
HHb
Cl⁻
Cl⁻
Hb⁻
O₂Hb⁻
K⁺
Na⁺
O₂
O₂
O₂

FIG. 33, 10A

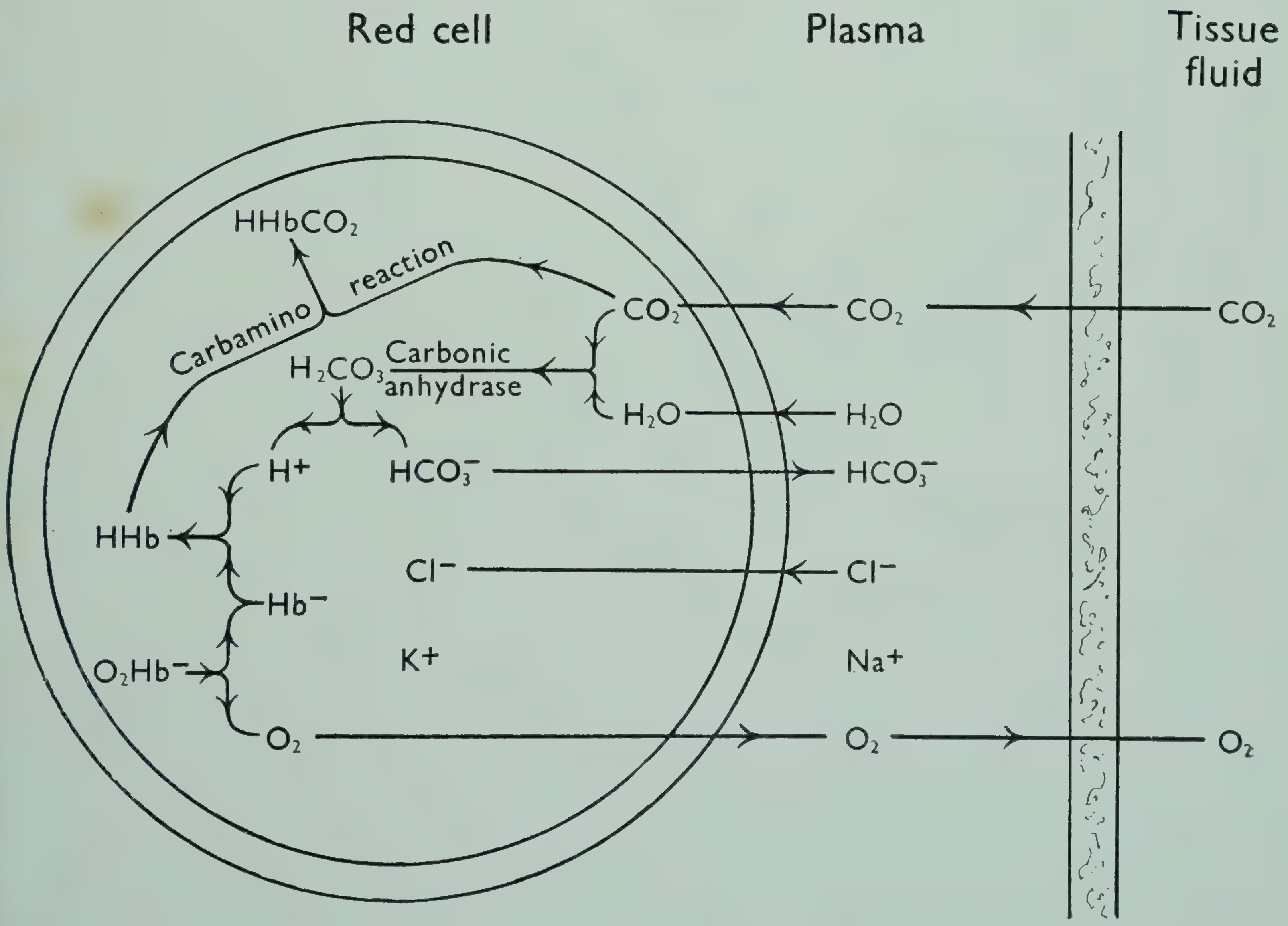

Red cell
Plasma
Tissue fluid
HHbCO₂
Carbamino reaction
CO₂
CO₂
CO₂
H₂CO₃
Carbonic anhydrase
H₂O
H₂O
H⁺
HCO₃⁻
HCO₃⁻
HHb
Cl⁻
Cl⁻
Hb⁻
O₂Hb⁻
K⁺
Na⁺
O₂
O₂
O₂

FIG. 33, 10B

A schematic summary of the chemical processes which occur when hæmoglobin takes up oxygen in the lungs (above) and parts with its oxygen in the tissues (below). The chloride shift is represented as a movement of ions. The shaded part represents the capillary wall. (*After* F. J. W. Roughton (1935), *Physiol. Rev.* **15**, 293.)

[*Facing page* 584

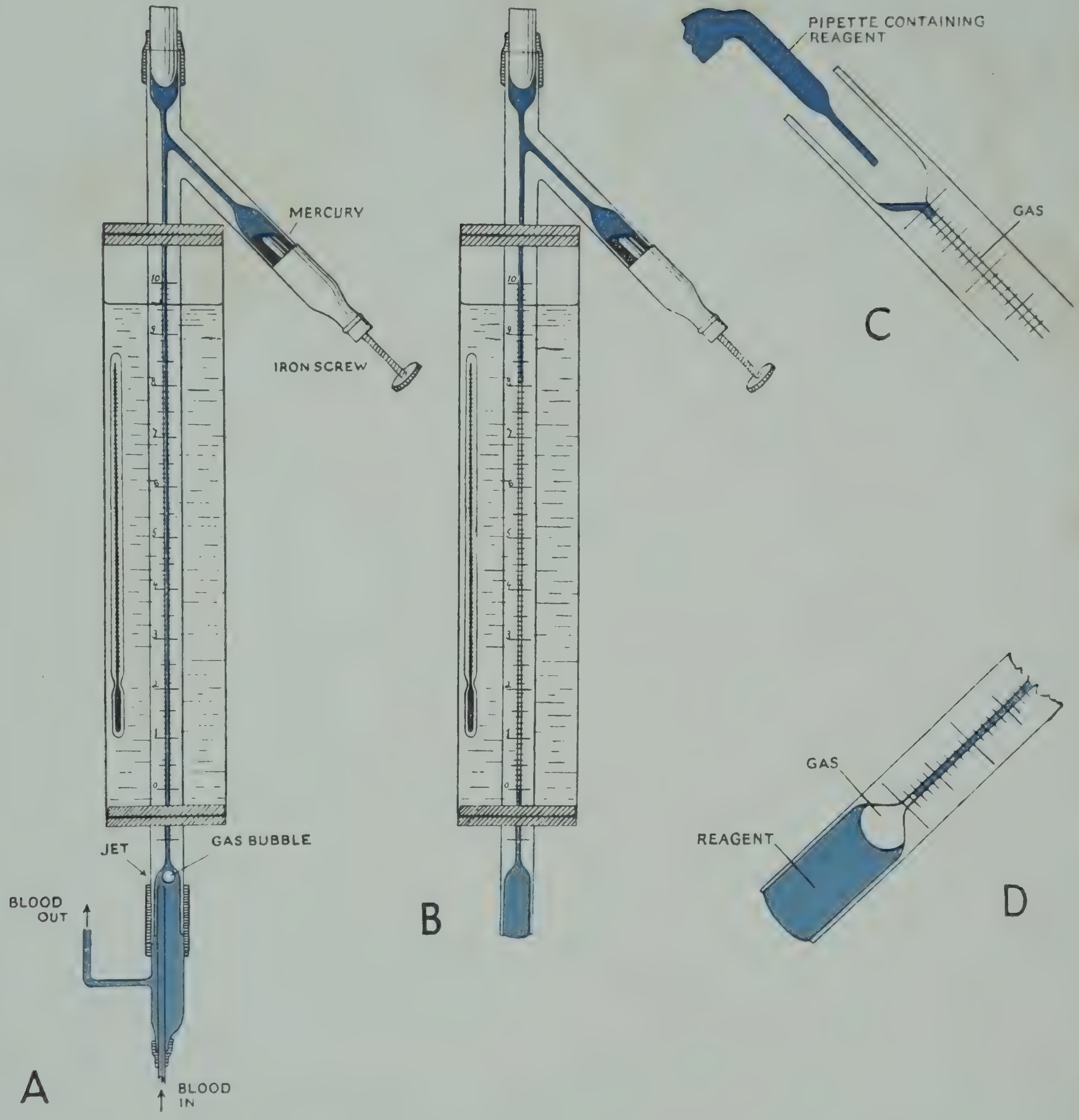

PIPETTE CONTAINING REAGENT
MERCURY
GAS
IRON SCREW
C
10
9
8
7
6
5
4
3
2
1
0
GAS
REAGENT
JET
GAS BUBBLE
BLOOD OUT
B
D
A
BLOOD IN

FIG. 33, 11

A. Krogh's microtonometer. Blood from an artery flows in at the lower end through a fine jet near the gas bubble which is kept oscillating by the incoming blood. The blood which emerges at the side arm can be re-injected into the animal. B. When equilibrium is reached between the gases in the blood and the gases in the bubble, the upper part of the apparatus is removed and by adjusting the screw the gas bubble is taken up into the graduated tube so that its volume is read off. C. The apparatus is then inverted and a solution containing potassium hydroxide or alkaline pyrogallol added with which the bubble is brought into contact, as at D. (Modified from A. Krogh (1908), *Skand. Arch. Physiol.* **20,** 266, 281, 282.)

perfused, that is to say has a high ventilation-perfusion ratio, the alveolar air approximates to dead space air and the R.Q. of the alveolar air sample is higher than that of the ideal alveolar air. By taking samples of alveolar air through catheters introduced into different parts of the lung it is possible to obtain information about the ventilation-perfusion relationships.

Measurement of the dead space (p. 549) gives a clue to the ventilation perfusion ratio since the space may be considered to include the anatomical dead space (where no gas exchange takes place) together with a contribution from alveoli with a high ventilation perfusion ratio (where less than average exchange occurs). If the dead space is more than 30 per cent. of the tidal volume a significant proportion of the alveoli are well ventilated but poorly perfused.

OXYGEN LACK

When the oxygen supply to the tissues becomes inadequate, anoxia or hypoxia is said to be present. Some aspects of this subject have already been discussed on p. 557. Although the effects upon the subject depend to some extent on the way in which the anoxia is produced, and by the rapidity with which it develops, the commonest effects are cerebral, namely, confusion, excitement, hallucinations, restlessness and unconsciousness. The causes of anoxia fall into four groups.

1. *Anoxic anoxia.*—This type of anoxia is due to defective oxygenation of the blood in the lungs; the arterial blood is not saturated to the usual value of 95 per cent. This is the commonest form of anoxia and is sometimes referred to as *anoxœmia.* It may occur when the exchange of oxygen between alveolar and air blood is impeded, for example by pulmonary œdema, collapse of the lung, or when the lung tissue is affected by pneumonia or emphysema. Rapid shallow breathing from any cause, or impaired air entry to the lung following obstruction to the air passages, or paralysis of the respiratory muscles all interfere with the oxygenation of the blood. Inspiration of air at high altitudes where the oxygen tension is low or of mixtures of anæsthetic gases with too low an oxygen content also produces anoxic anoxia. Finally this kind of anoxia occurs in certain congenital cardiac defects, in which large amounts of blood are shunted from the right to the left side of the heart so that well oxygenated blood in the aorta is mixed with poorly oxygenated blood that has not traversed the pulmonary circuit.

In all these circumstances oxygen is supplied to the tissues of the body at a much lower tension than normal and may be inadequate for their needs, especially during activity. Anoxæmia stimulates respiration through the carotid and aortic bodies; dyspnœa is produced and the elimination of carbon dioxide is increased. The rapid shallow breathing often seen in this condition aggravates the anoxia and alkalosis (p. 643) further reduces the uptake of oxygen by the tissues (Fig. 33, 4). If the condition is not relieved, peripheral circulatory failure (p. 524) develops and anoxia of the stagnant type, described below, further reduces the supply of oxygen to the tissues.

Anoxic anoxæmia is associated with a high cardiac output, an abnormally short circulation time, an increased pulse pressure and dilatation of tbe peripheral arterioles—the so-called hyperkinetic circulation. These phenomena all have the effect of increasing the oxygen supply to the peripheral tissues.

2. *Anæmic anoxia.*—This results from interference with oxygen transport either because there is a reduction in the amount of circulating hæmoglobin (anæmia), or because the hæmoglobin is rendered unavailable by reason of its being combined with any substance with which it forms a stable compound. In this type of anoxia the oxygen load, that is the oxygen carried per 100 ml. of blood, is reduced but, since the oxygen tension in the alveolar air and in the arterial blood is normal, the tissues receive their oxygen at a normal tension. As the blood flows through the capillaries in the tissues the oxygen tension falls very rapidly, so that if no compensation occurred the greater part of the oxygen would be supplied to the tissues at a much reduced tension. In anæmia, however, this disadvantage is compensated for by an increase in the cardiac output; the circulation rate is more rapid, and each 100 ml. of blood is required to part with less of its oxygen load to the tissues than would otherwise be the case. Oxygen is thus supplied to the tissues at tensions higher than could occur with the normal circulation rate. As a result of this compensatory mechanism anæmic patients are rarely breathless at rest, although they may rapidly become so with exercise.

Hæmoglobin may be rendered unavailable for oxygen transport by drugs such as chlorates and nitrates which convert it to methæmoglobin, or by carbon monoxide which has an affinity for hæmoglobin about 300 times greater than that of oxygen. For this reason concentrations of carbon monoxide as low as 0·1 per cent. in the atmosphere produce dangerous anoxia; where there is prolonged exposure, 0·01 per cent. is the maximum concentration which can be tolerated. In carbon monoxide poisoning the oxygen dissociation curve is shifted to the left, so that the blood cannot give off its oxygen to the tissues unless their oxygen tension is extremely low.

It was shown thirty years ago by experiments in which animals were poisoned by carbon monoxide that the quickest way to achieve recovery was to make the animals breathe a mixture of carbon dioxide and oxygen. On this basis the administration of 5 per cent. carbon dioxide in oxygen was for long advised in resuscitation, especially in cases of coal gas poisoning. In 1952 however a Committee of the Medical Research Council reconsidered the subject and recommended that oxygen only should be used in resuscitation. The victim of severe lack of oxygen is unconscious and his respiratory centre is depressed and in spite of the mounting carbon dioxide content of his tissues he does not breathe; administration of carbon dioxide can only make his condition worse. The cells of the brain must have oxygen quickly if they are to survive and this can best be achieved by administering oxygen and applying artificial respiration. Even if an anoxic patient is breathing his respiratory centre may be so depressed that it cannot, as in a normal person, prevent a rise in the carbon dioxide content of the body when more carbon dioxide is administered. Carbon dioxide in high enough concentrations is an anæsthetic and delay in recovery from surgical anæsthesia may be due to retention of carbon dioxide. In certain acute and chronic respiratory diseases carbon dioxide retention may result in unconsciousness and depression of respiration in spite of efficient administration of oxygen. To administer additional carbon dioxide to such patient would be dangerous or even lethal. Under emergency conditions it is seldom possible to make an accurate diagnosis of the cause of respiratory failure. Since the disadvantages of administering carbon dioxide are overwhelmingly greater than its advantages resuscitation should be carried out, unless in ideal conditions in a hospital, with oxygen only.

3. *Stagnant anoxia.*—As its name implies, this type of anoxia is the result of slowing of the circulation, such as may occur locally as the result of obstruction to the venous outflow from a part, or more generally because of cardiac failure or peripheral circulatory failure (shock). In this condition arterial oxygen saturation, the partial pressure of oxygen and the oxygen load are all normal, but the circulation rate is slowed and the blood dwells longer in the capillaries. The result is that the arteriovenous oxygen difference is increased and much of the oxygen is delivered to the tissues at a reduced tension. A small degree of compensation can, however, occur since the retarded flow leads to increase in the capillary blood carbon dioxide tension which encourages the dissociation of oxyhæmoglobin and thus increases the supply of oxygen to the tissues.

4. *Histotoxic anoxia.*—The only important example of histotoxic anoxia is cyanide poisoning which, by paralysing the cytochrome system and other iron-containing enzyme systems, interferes with tissue oxidation. The cells are, therefore, unable to utilize the oxygen conveyed to them and the arteriovenous oxygen difference is much reduced. Narcotics may have a similar effect by interfering with dehydrogenase systems.

CYANOSIS

As we have already seen, the colour of the skin is determined by the amount and condition of the blood in the minute vessels and, especially, in the veins of the subpapillary plexus (p. 519). *Cyanosis* is a violet or sometimes greyish colour of the skin produced by abnormally large amounts of reduced hæmoglobin, or, less frequently, by abnormal pigments such as methæmoglobin and sulphæmoglobin. Cyanosis depends on the absolute amount of reduced or abnormal hæmoglobin in the blood and is not present until at least 5 g. per 100 ml. of hæmoglobin is in the reduced state. A normal subject, then, with 14·5 g. per cent. does not appear cyanosed until about one-third of the circulating hæmoglobin is unsaturated; in other words the oxygen saturation of the blood in the subpapillary venous plexus must be less than 65 per cent. It follows that an anæmic person with a total hæmoglobin concentration of less than 5 g. per 100 ml. cannot appear cyanosed even if he is suffering from hypoxia sufficient to produce irreparable damage to the central nervous system.

If we assume that normal blood contains 15 g. Hb per 100 ml. with an oxygen capacity of 20 ml. per 100 ml. of blood, and that 100 ml. mixed venous blood contains 14 ml. oxygen, we can calculate that venous blood contains $\frac{20-14}{20} \times 15 = 4{\cdot}5$ g. hæmoglobin in the reduced state. Since a capillary contains arterial blood at one end and venous (reduced) blood at the other the *average* amount of reduced hæmoglobin is likely to be much less than 4·5 g. per 100 ml. and accordingly the skin does not normally appear cyanosed. When the oxygen saturation of the arterial blood is reduced (anoxic anoxia, p. 585) cyanosis tends to occur very readily, much more readily than with stagnant anoxia where the arterial blood is fully saturated.

The cyanosis sometimes seen in persons with polycythæmia in whom the concentration of red cells is over eight million per cu. mm and of hæmoglobin over 23 g. per 100 ml. is mainly the result of stagnant anoxia produced by the increased viscosity imparted to the blood by its much greater cellular content.

OXYGEN THERAPY

The administration of oxygen often benefits strikingly certain kinds of anoxia and relieves cyanosis. When the arterial oxygen saturation is normal, as in the stagnant, anæmic and histotoxic forms of anoxia, less is to be expected of oxygen therapy than in the anoxic type, in which the arterial oxygen saturation is reduced. Although the efficient administration of oxygen (Fig. 33, 12) can raise the partial pressure of this gas in the alveolar air from 100 to 600 mm Hg it does not necessarily follow that the arterial oxygen tension is thereby increased. However, in the anoxæmia due to paralysis of the muscles of respiration, to rapid shallow breathing and to many diseases of the lung in which a considerable proportion of the pulmonary blood flow perfuses areas that are not well aerated, the administration of oxygen relieves cyanosis rapidly and the effects of anoxia more slowly. Such a dramatic relief is not to be expected when there is a large shunt of venous blood from the right to the left side of the heart, or when the diseased pulmonary tissue has been largely deprived of its blood, because in these conditions the blood leaving the functioning alveoli is already 95 per cent. saturated.

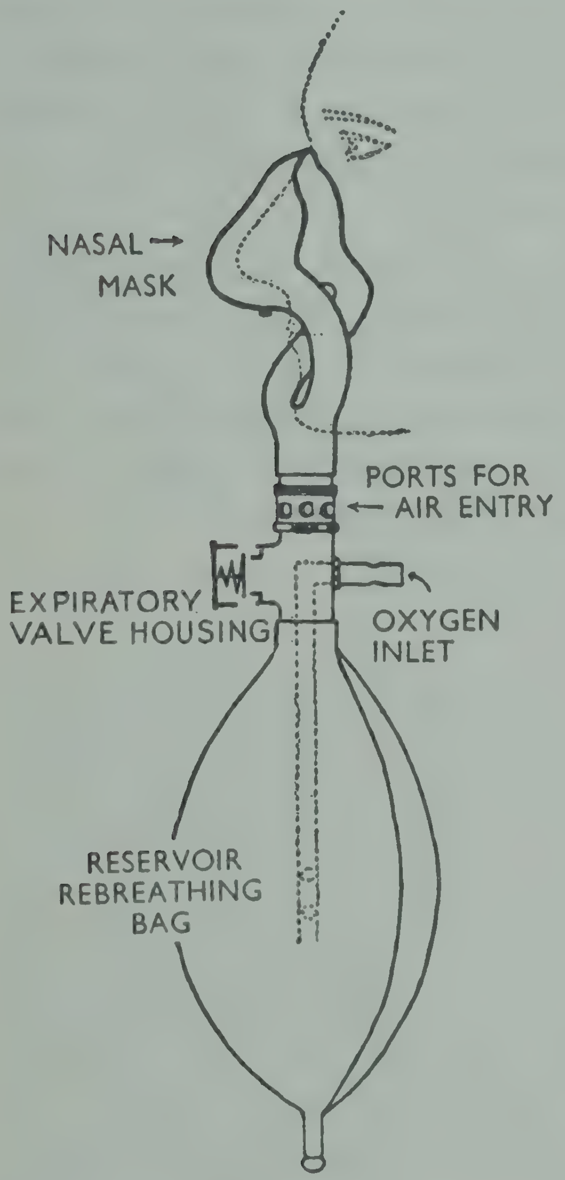


Fig. 33, 12

The B.L.B. inhalation apparatus for administration of oxygen (W. I. Card, J. F. Smith, W. J. Griffiths, B. A. McSwiney & B. Savage (1940), *Lancet*, i, 99).

Efficient oxygen administration at atmospheric pressure is capable of increasing the concentration of oxygen in solution in the plasma from the normal level of about 0·3 vol. per 100 ml. to values of the order of 1·8 vol. per 100 ml., which is quite a significant increase. The normal arteriovenous oxygen difference is about 5 or 6 ml. per 100 ml., and this is increased in anoxia. Administration of oxygen may, therefore, by increasing the plasma concentration, make an important contribution to the oxygen supply of the tissues, equivalent perhaps to an increase of 30 per cent. in cardiac output. For this reason the administration of oxygen is useful in all kinds of anoxia, however produced; its value can be judged in practice by observing the effect in dispelling cyanosis.

Pure oxygen may be administered through a catheter inserted through the nostrils into the nasopharynx. At a flow of four litres per minute the oxygen tension of the inspired air is more than doubled and higher tensions can be obtained by using two catheters. The oxygen supplied during expiration is partly wasted. It is more economical to use a face mask, one type of which is shown in Fig. 33, 12. This mask, covering the nose only, is connected to a system of ports which allow for the dilution of the inspired oxygen with atmospheric air and during expiration the oxygen, which is being supplied continuously, collects in a rubber bag. By far the most efficient and comfortable apparatus for oxygen administration is the oxygen tent, which is constructed of flexible plastic material and fits over the head and trunk of the patient, who can move about fairly freely within. The oxygen concentration of the atmosphere in the tent depends on the rate of delivery of the gas and on the efficiency of the apparatus, but there is little difficulty in maintaining oxygen concentrations of 40 to 60 per cent. in the tent over many days. It is necessary to keep the enclosed atmosphere cool by chilling the entering gas and to make arrangements to absorb carbon dioxide.

The value of oxygen inhalation in flying at high altitudes was demonstrated in the 1939–45 War; the signs of oxygen lack have already been described on pp. 557 and 585. It is essential to breathe through an oxygen mask when flying at altitudes above 15,000 feet (5000 metres). At great heights when the barometer reads less than 150 mm Hg even pure oxygen must be administered under positive pressure; it is the pressure (or partial pressure) of oxygen rather than its concentration that determines the arterial oxygen saturation.

REFERENCES

Armstrong, H. G. (1952). Principles and practice of aviation medicine. 3rd ed. London: Baillière, Tindall & Cox.

Arnott, W. M. (1955). Order and disorder in pulmonary function. *Brit. med. J.* ii, 279–284 and 342–348.

Barcroft, J. (1928). *Respiratory Function of the Blood.* Cambridge University Press.

Collected Papers on Aviation Medicine (1955). Advisory group for aeronautical research and development. London: Butterworth.

Comroe, J. H. & Dripps, D. R. (1950). *The Physiological Basis for Oxygen Therapy.* Oxford: Blackwell.

Donald, K. W. (1953). The definition and assessment of respiratory function. *Brit. med. J.* i, 415–422 and 473–478.

Peters, J. P. & Van Slyke, D. D. (1946). *Quantitative Clinical Chemistry.* Vol. I. 2nd ed. London: Baillière, Tindall & Cox.

Rahn, H. (1949). A concept of mean alveolar air and the ventilation blood flow relationships during pulmonary gas exchange. *Amer. J. Physiol.* **158,** 21–30.

Riley, R. L. & Cournand, A. (1949). Ideal alveolar air and the analysis of ventilation-perfusion relationships in the lungs. *J. appl. Physiol.* **1,** 825–847.

Roughton, F. J. W. (1935). Recent work on carbon dioxide transport by the blood. *Physiol. Rev.* **15,** 241–296.

Roughton, F. J. W. & Kendrew, J. C. (Editors) (1949). *Hæmoglobin—a Symposium in memory of Sir Joseph Barcroft.* London: Butterworth.

Stewart, C. P. & Dunlop, D. M. (1954). *Clinical Chemistry in Practical Medicine.* 4th ed. Edinburgh: Livingstone.

CHAPTER 34

THE KIDNEY

THE kidneys are much more than excretors of waste products. Although the removal from the body of the end-products of nitrogen metabolism by their elimination in the urine is important, and indeed essential to life, the less obvious role of the kidney in regulating the composition of the blood plasma, its homoiostatic function, is now recognized as being one of the kidney's major activities.

STRUCTURE OF THE KIDNEY

The kidneys lie in the posterior part of the abdominal cavity on either side of the vertebral column and each is supplied by a short renal artery which arises directly from the aorta. The arteries enter and the veins leave at the hilum; the pelvis of the ureter, the funnel-shaped upper end of the ureter formed by the joining together of about ten small branches or *calcyes*, also emerges from the kidney at the hilum. The kidney proper consists of an external reddish-brown *cortex* and an internal pale *medulla* made up of *pyramids*, the apices or papillæ of which dip into the calyces (Fig. 34, 2).

The minute anatomy of the kidney is exceedingly difficult to understand from the ordinary histological section because it consists of an enormous number of intertwined tubules. The first accurate description was given by Bowman in 1842. He injected saturated solutions of potassium dichromate and of lead acetate in succession into the renal artery and produced a deposit of lead chromate in the glomeruli; occasionally, when the pressure of injection was high, the capillaries ruptured and lead chromate found its way into the glomerular capsule (Bowman's capsule) and into the convoluted tubule (see Fig. 34, 3). The kidney unit is a tubule about 3 cm long called a *nephron* (Fig. 34, 1), which begins as a blind dilated end (*glomerular capsule*) and then pursues a very tortuous course through the cortical and medullary substances, finally opening into a collecting tubule. The glomerular capsule can be thought of as a sphere, lined by epithelium lying on a basement membrane, invaginated at one place by a tuft or *glomerulus* of capillary blood vessels of which the afferent vessel is larger than the efferent. A glomerulus surrounded by its glomerular capsule is called a *Malpighian* or *renal corpuscle*, each of which is about 0·2 mm in diameter. There are about a million of these bodies in each human kidney and it has been calculated that the total surface area of all the glomerular capillaries of both kidneys exceeds 1 sq. metre. The wall of a glomerular capsule is made of flattened cells, which change fairly abruptly to cubical cells where the neck of the capsule becomes the first segment of the tubule at a point opposite the entry of the glomerulus (Fig. 34, 4, *a*). The short neck is continued into the *first* (or proximal) *convoluted tubule* and into the descending limb of the loop of Henle, which passes down into the medulla; the tube then turns back to form Henle's loop and re-enters the cortical substance to form the *second convoluted tubule*. This last, becoming a *junctional tubule*, joins a *collecting tubule* which finally opens into a calyx. The types of cells lining the tubules vary in different parts of the nephron. The epithelium of the convoluted tubules and of the ascending limb of the loop of Henle is cylindrical and shows granules arranged radially (rodded epithelium). The cells of the first convoluted tubule have a striated free border (brush border) (Fig. 34, 4, *b*) not unlike that seen on the cells covering the intestinal villi. The descending limb of the loop of Henle has a flattened epithelium, while the junctional and collecting tubules can be distinguished by their clear cubical or columnar cells.

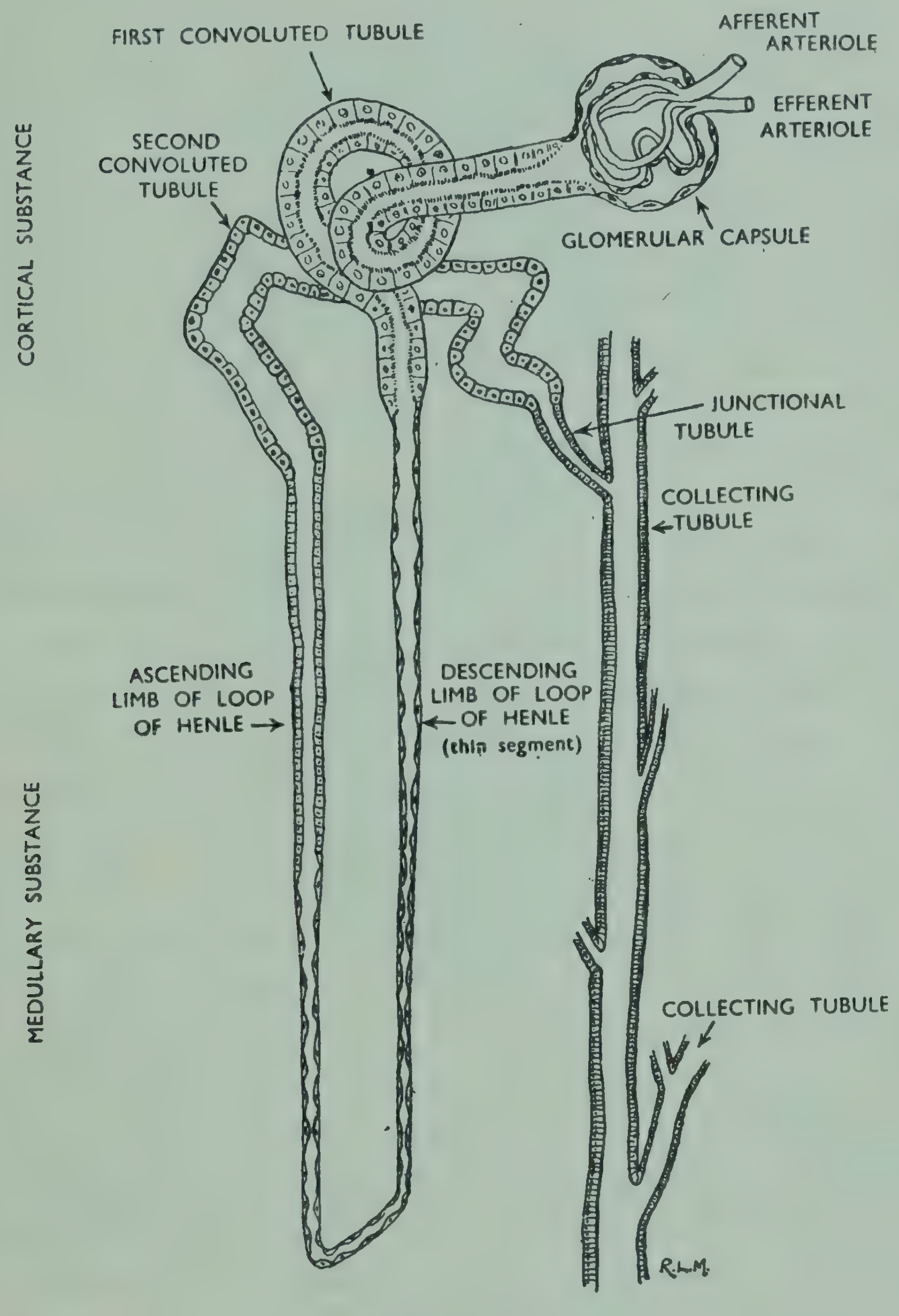

FIRST CONVOLUTED TUBULE
AFFERENT ARTERIOLE
EFFERENT ARTERIOLE
CORTICAL SUBSTANCE
SECOND CONVOLUTED TUBULE
GLOMERULAR CAPSULE
JUNCTIONAL TUBULE
COLLECTING TUBULE
ASCENDING LIMB OF LOOP OF HENLE
DESCENDING LIMB OF LOOP OF HENLE (thin segment)
MEDULLARY SUBSTANCE
COLLECTING TUBULE
R.L.M.

Fig. 34, 1

The nephron and collecting tubule diagrammatically represented. (*After* Homer Smith.)

The main divisions of the renal artery within the kidney, called *arciform arteries*, run between cortex and medulla parallel to the surface of the organ. At frequent intervals branches arise from them and run in the cortex at right angles to the surface. From each of these vessels numerous afferent glomerular arterioles arise (Fig. 34, 4, *c*) which almost at once break up into the capillaries of the glomerular tuft. Blood leaves the glomerulus by the efferent glomerular arteriole, notably smaller in calibre than the afferent arteriole, and this immediately breaks up into a second capillary network (Fig. 34, 4, *d*) surrounding the convoluted tubules in the cortex from which the blood eventually passes into the renal veins.

Although this is the arrangement throughout the cortex both in animals and in man, the glomeruli situated in the deeper layers of the cortex (*juxta-medullary glomeruli*) are somewhat larger in size and have a different plan. Arising in many cases directly from a large artery the afferent glomerular arteriole breaks up into the glomerular capillaries which are then drained by an efferent vessel at least as large in calibre as the afferent arteriole. After a short straight course towards the medulla this vessel breaks up into a number of long straight vessels which follow the course of the loops of Henle before draining into the renal veins (Fig. 34, 4, *e*). It has recently been shown, both in the rabbit and in the rat, that a large proportion of the total renal blood flow can be shunted through the medullary vascular bed leaving the cortex relatively ischæmic; at present, however, no physiological function can be assigned to this phenomenon. Tests in man provide no evidence for such a shunt mechanism because the mass of functioning renal tissue remains relatively constant under widely different experimental circumstances.

Renal Blood Vessels.—There is normally no anastomosis between branches of the renal artery although numerous communications exist between the intertubular capillaries, and arteriovenous anastomoses analogous to those present in the skin (p. 207) are also present in the kidney. All the blood issuing from the renal veins has perfused two capillary beds, namely the glomerular tuft and the peritubular capillaries, since the direct supply of blood to tubules from the renal artery is insignificant.

It has already been stated that the kidneys are supplied with blood by two short renal arteries arising directly from the abdominal aorta; the kidneys are therefore perfused with blood at the highest available pressure. In a normal individual the renal blood flow amounts to some 1200 ml. per minute, or as much as one-quarter of the cardiac output. Arrest of the renal circulation could increase the total peripheral resistance by about one-third and vasoconstriction in the renal beds could make large amounts of blood available to other structures, as for example in peripheral circulatory failure (p. 524) and when the effective circulating blood volume is reduced. However, experimental evidence indicates that the renal vessels are not involved in vasomotor reflexes.

The Renal Nerves.—The nerve supply of the kidney is mainly sympathetic, chiefly from the tenth, eleventh and twelfth thoracic segments of the spinal cord, but a small and variable number of parasympathetic fibres is derived from the vagus. The renal nerves run in a plexus along the renal arteries to supply the blood vessels and, according to some authors, the renal tissue proper. Experimental stimulation of the nerves to the kidney produces vasoconstriction and a reduction in urinary output; there is no evidence that either glomerular or tubular activity is affected directly. In point of fact the denervated kidney is able to respond to changes in the composition of the blood almost as effectively as the healthy innervated organ. In the normal unanæsthetized animal denervation of the kidney does not increase the flow of urine, and in man section of the renal nerves has only slight and inconstant effects on renal blood flow.

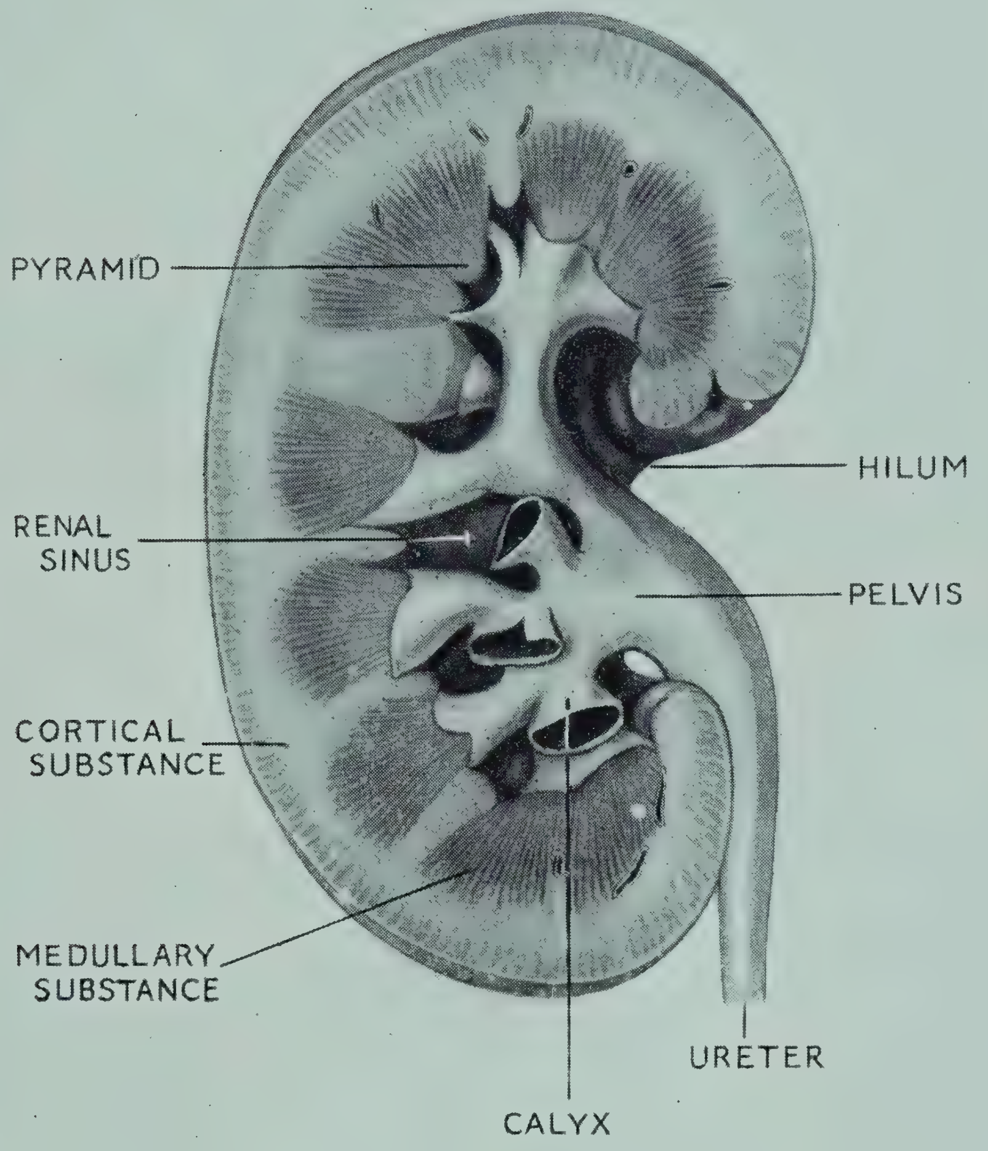

PYRAMID
HILUM
RENAL SINUS
PELVIS
CORTICAL SUBSTANCE
MEDULLARY SUBSTANCE
URETER
CALYX

Fig. 34, 2

A vertical section through a kidney to show the main features of its structure.

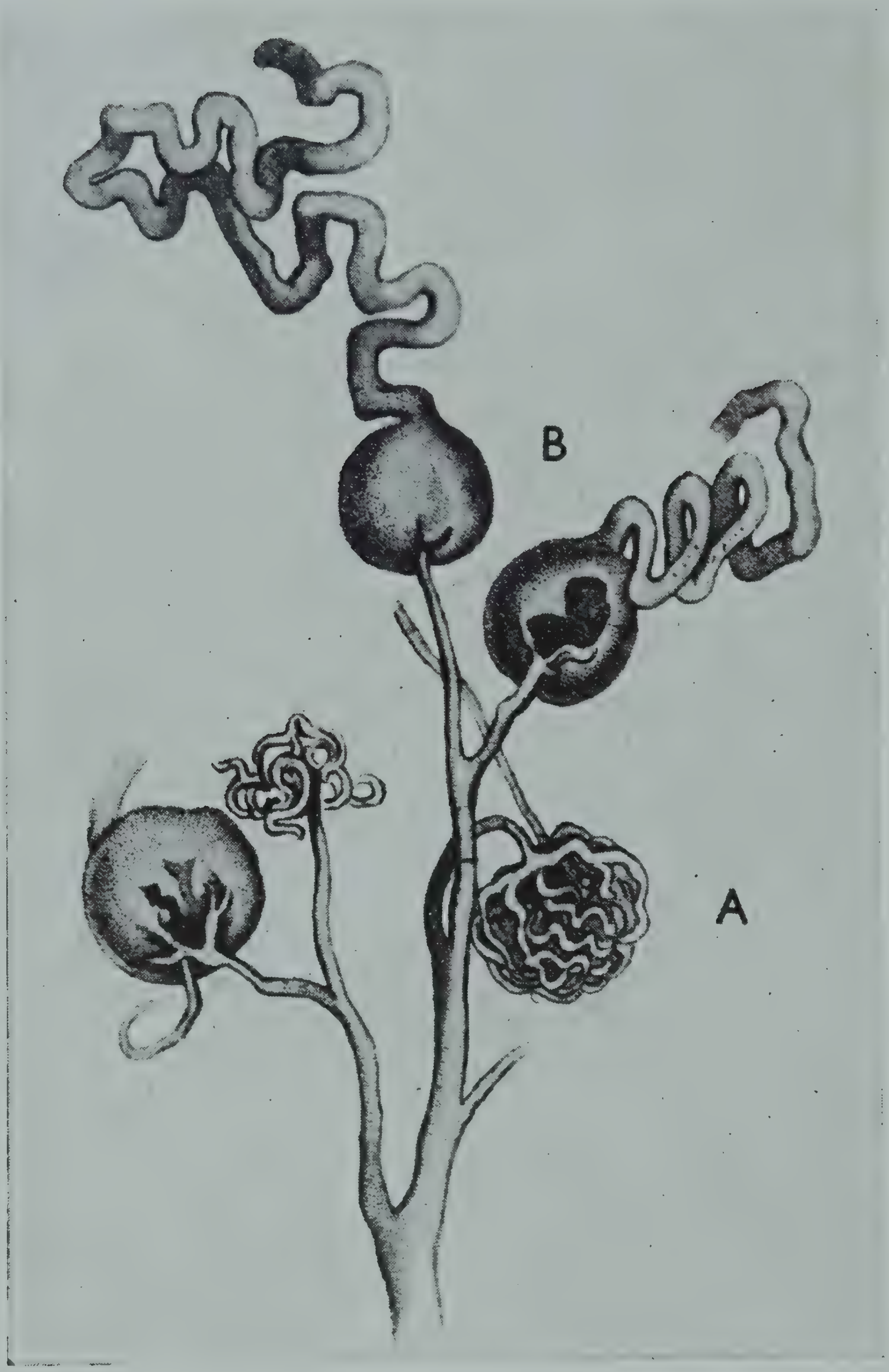


Fig. 34, 3

At A the tuft of capillaries (the glomerulus) is shown. At B the capillaries have given way and the lead chromate has passed into Bowman's capsule and into the first part of the renal tubule (W. Bowman (1842), On the structure and use of Malpighian bodies of the kidney, *Phil. Trans.* **132**, Fig. 10, p. 79).

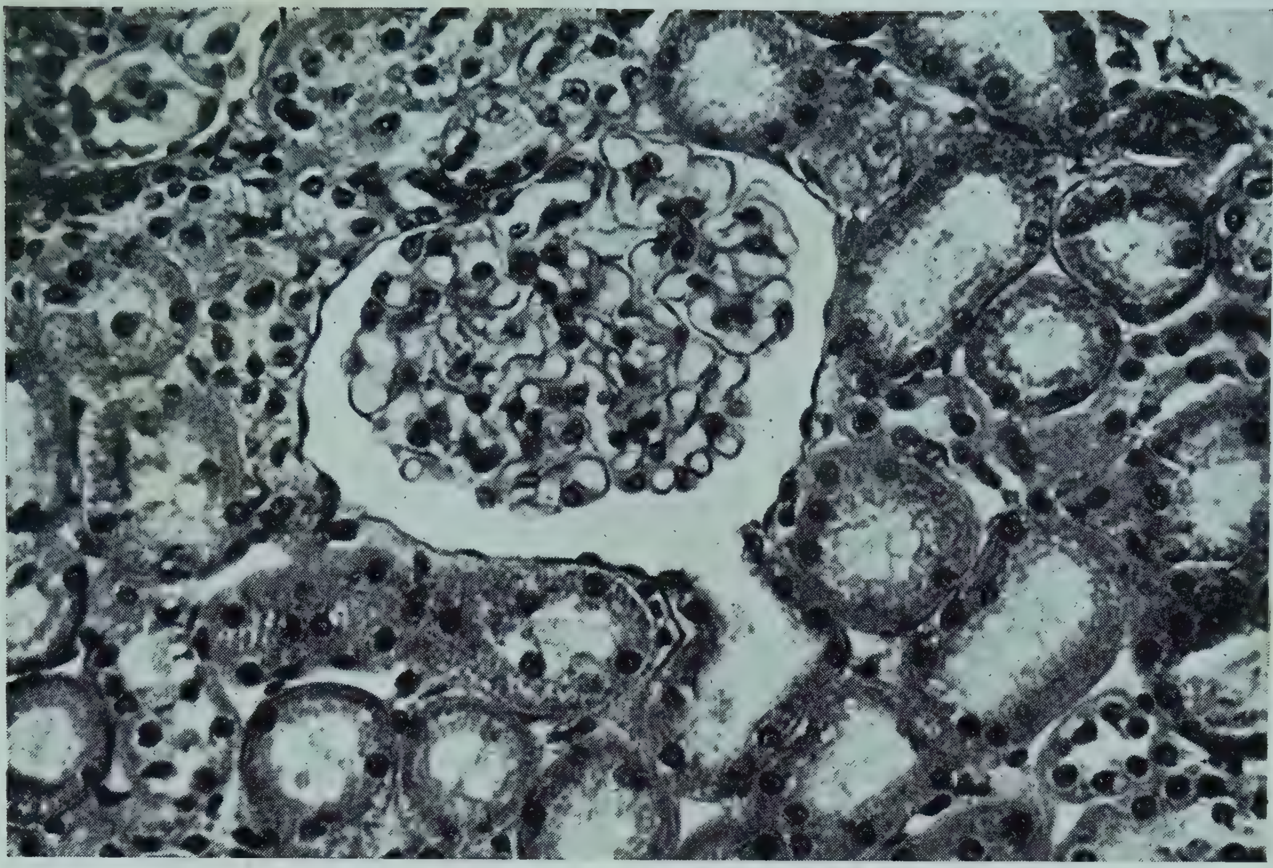


FIG. 34, 4, *a*

Photomicrograph of rat kidney to show glomerular capsule composed of flat cells and the first convoluted tubule emerging from it.

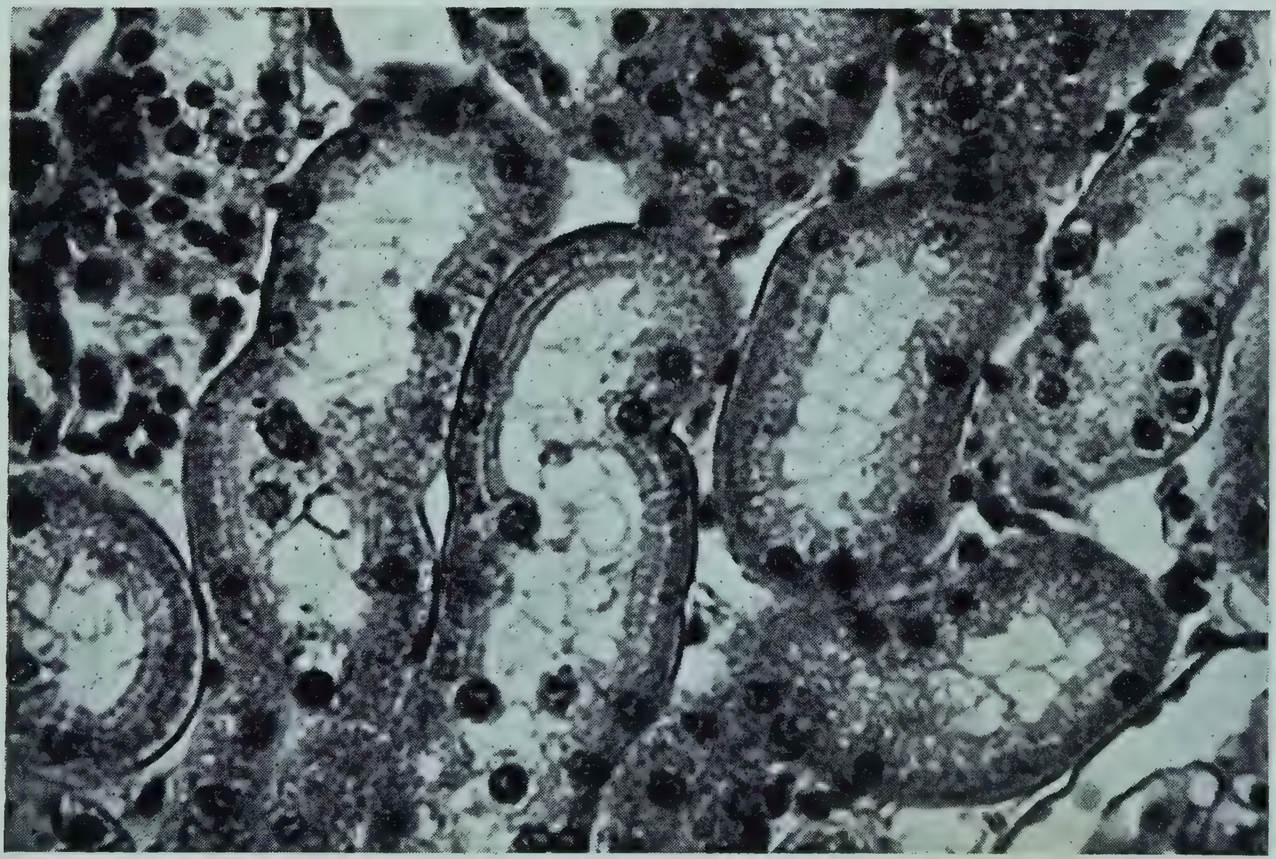


FIG. 34, 4, *b*

Section through the first convoluted tubule to show rodded epithelium with 'brush border.'

(*By courtesy of W. W. Park.*)

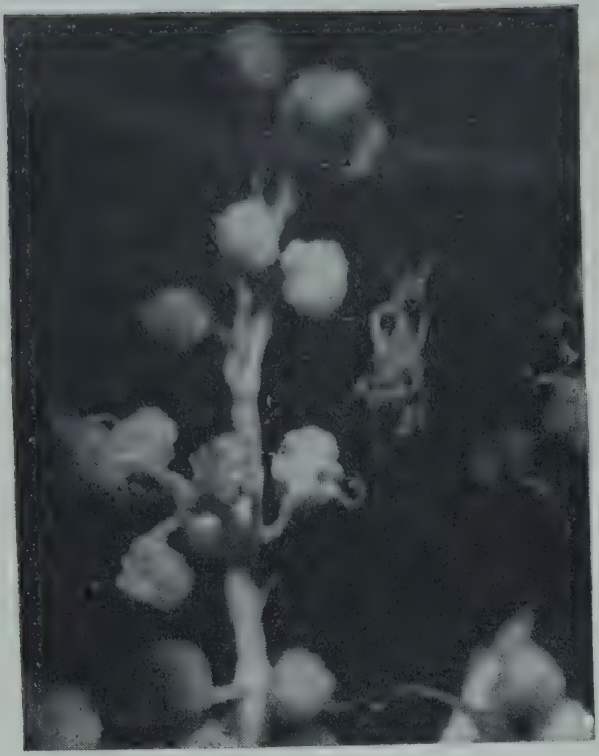


Fig. 34, 4, *c*

Micro-dissection of a branch from an arciform artery, showing afferent glomerular arterioles and glomerular tufts after injection of rubber latex in a rabbit (J. Trueta, A. E. Barclay, K. J. Franklin, P. M. Daniel & M. M. L. Prichard (1947), *Studies of the Renal Circulation*, p. 57. Oxford: Blackwell Scientific Publications).

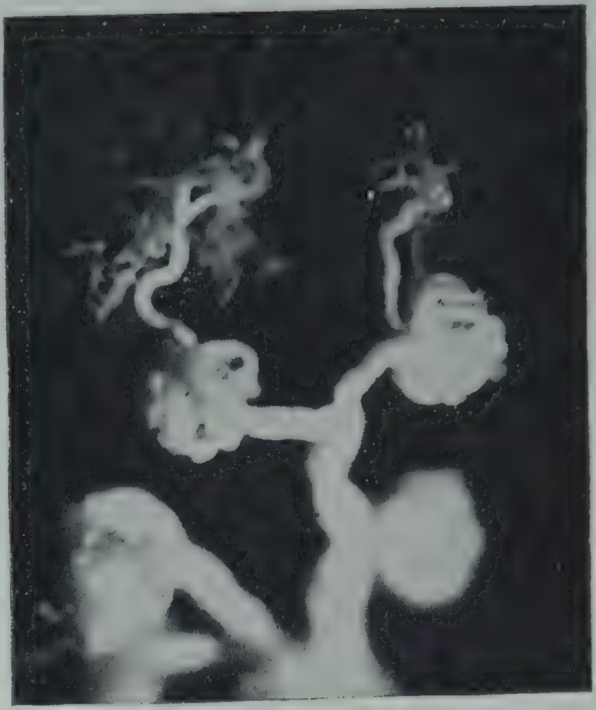


Fig. 34, 4, *d*

Cortical glomerulus in man dissected from a kidney injected with rubber latex. Note large afferent and smaller efferent arteriole, the latter breaking up into a network of capillaries (Trueta *et al.*, *Studies of the Renal Circulation*, p. 69. Oxford: Blackwell Scientific Publications).

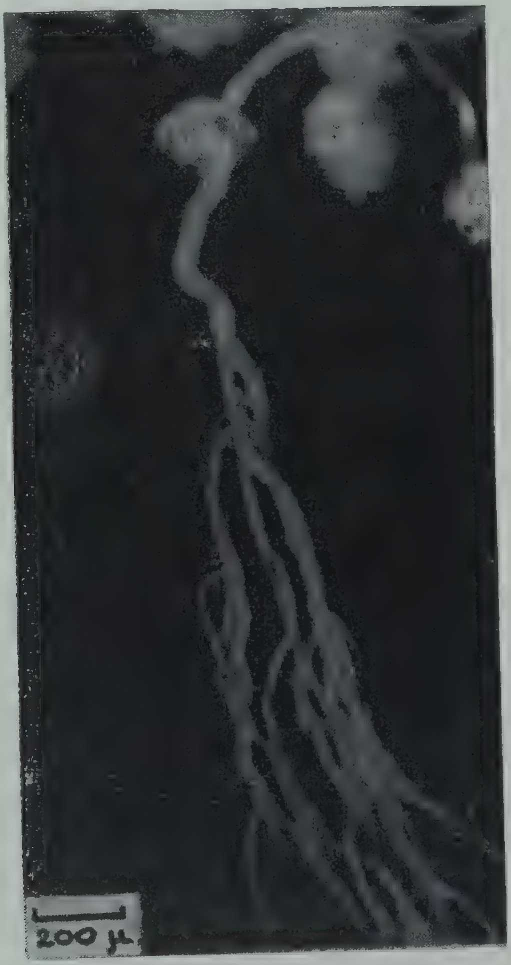


Fig. 34, 4, *e*

Juxta-medullary glomerulus in man dissected from a kidney injected with rubber latex. Note that afferent and efferent arterioles are of equal size (Trueta *et al.*, *Studies of the Renal Circulation*, p. 72. Oxford: Blackwell Scientific Publications).

GLOMERULAR FUNCTION

The starting-points from which we must consider the problem of the physiology of the kidney are the structure of the nephron and a comparison of the composition of the fluid reaching the kidney (blood plasma) and the fluid leaving it, that is the urine (Table 34, 5). We have, for example, to account for the absence of protein and glucose in urine, for the very high concentration in urine of ammonia and creatinine when other substances such as sodium and calcium may appear in nearly the same concentration in urine and plasma.

Table 34, 5

Comparison of Representative Values for the Concentration of Certain Substances in Plasma and Urine of Man

(The values given in this table are approximate. The range of variation in healthy persons is considerable.)

	Plasma g. per 100 ml.	Urine g. per 100 ml.	Ratio $\frac{\text{Urinary conc.}}{\text{Plasma conc.}}$
Water	90 to 93	95	..
Proteins and other colloids .	7 to 8	0 *	..
Urea	0·03	2	60
Uric acid	0·002	0·03	15
Glucose	0·1	0·0 *	..
Creatinine	0·001	0·1	100
Ammonia	0·0001	0·05	500
Sodium	0·32	0·6	2
Potassium	0·02	0·15	7
Calcium	0·01	0·015	1·5
Magnesium	0·0025	0·01	4
Chloride, as Cl.	0·37	0·6	2
Phosphate, inorganic, as P .	0·003	0·12	40
Sulphate, as H_2SO_4 . . .	0·003	0·18	60

* The usual tests are negative but more delicate tests show that traces of glucose and protein are present in normal urine.

Both Ludwig and Bowman in the middle of last century guessed that the function of the glomerular capsule was to produce a filtrate of blood plasma, but they did not agree about the role of the tubules. The renal corpuscle seems designed for filtration and receives its blood almost directly from the aorta. Thus the hydrostatic pressure in the glomerulus, being greater than the osmotic pressure of the plasma proteins (see pp. 369 to 373), is sufficient to cause water and simple solutes of low molecular weight to pass through the endothelium of the capillaries and through the glomerular capsule. The latter is considered by some histologists to be an incomplete network of cells, the *pericytes*, so that the glomerular filtrate has to pass through a single layer of cells, the capillary endothelium, a layer of basement membrane and then between the cells of the supporting pericytic reticulum to reach the interior of the glomerular capsule.

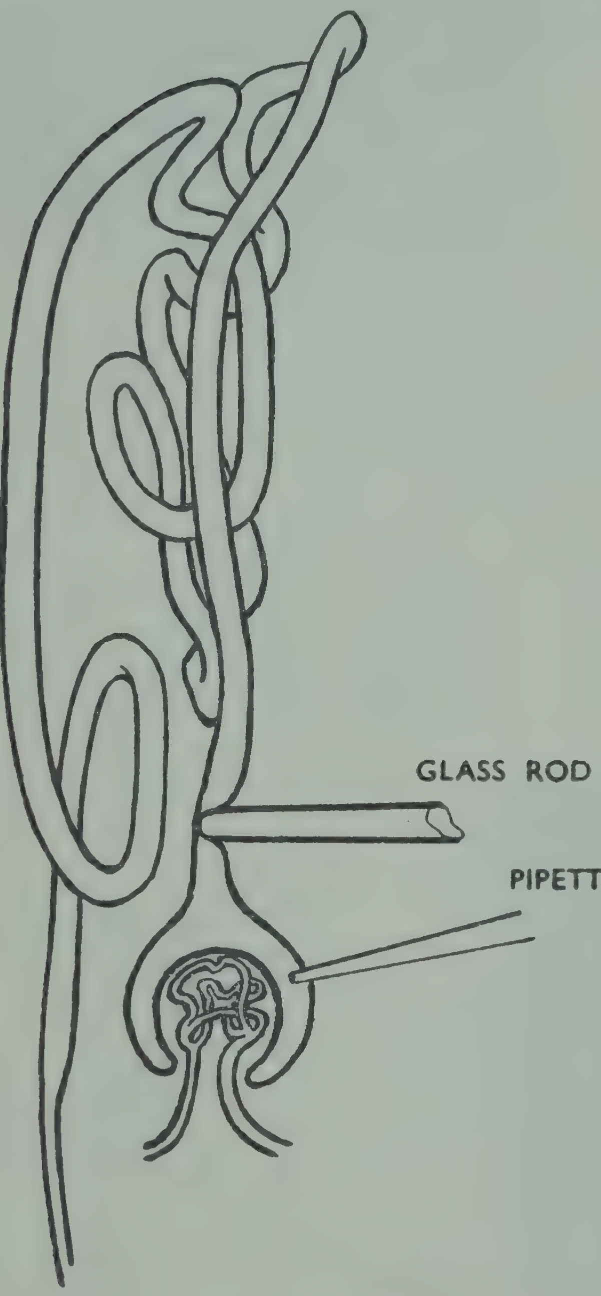


Fig. 34, 6

Method of obtaining glomerular fluid by a pipette (10 μ in diameter) inserted into the glomerular capsule. The tubule is shut off by pressure with a blunt-ended glass rod.

In the last twenty-five years methods have been evolved for obtaining glomerular fluid from frog kidneys; it is unfortunately much more difficult, although not impossible, to obtain samples from mammalian kidneys because the glomeruli only rarely lie on the surface of the kidney. By the use of a micromanipulator, a mechanical apparatus for carrying out very small movements, a tiny pipette can be inserted into a glomerular capsule and a specimen of glomerular filtrate obtained for analysis (Fig. 34, 6). The studies of Richards, Walker and their colleagues have shown that the glomerular fluid consists essentially of plasma without protein; it contains the same concentration of chloride, glucose and urea as blood plasma. The polysaccharide inulin, after injection into the frog's circulation, is also found in the same concentration in the plasma and in the glomerular filtrate. This relatively large molecule (molecular weight 5200) is thus transferred across the glomerular membrane in the same way as urea and chloride are. This transference cannot be a diffusion process, because in that case the large molecules would pass across much more slowly than the small molecules. The only possible explanation of the facts is that the process is one of filtration in which all molecules smaller than the pores of the filter pass through it so that the only difference between the fluid to be filtered (plasma) and the filtrate is the absence in the latter of substances of large molecular size. Clearance studies with dextran, another polysaccharide of which solutions can be prepared containing molecules of different sizes, have shown that the pores in the glomerular membrane are not of equal sizes. All are permeable to dextran molecules of molecular weight 5000; 20 per cent. pass molecules of 30,000 and a few allow molecules of molecular weight up to 70,000 to pass.

Before these direct methods became available there was indirect evidence for glomerular filtration and it was realized that it was the function of the tubules to modify the glomerular filtrate either by absorption or by secretion. Thus, if an excised dog's kidney is perfused with warm oxygenated blood from a heart-lung preparation the urine it secretes contains less than 0·2 per cent. sodium chloride as compared with 0·65 per cent. in the plasma, while the concentration of creatinine in the urine exceeds by two or three times its concentration in the plasma. If now the temperature of the kidney is lowered to 25° C, the urine contains more chloride and much less creatinine, approaching in fact an ultrafiltrate of blood plasma. The explanation is that the physical process of glomerular filtration is little affected by the reduction in temperature, whereas the metabolic rate of the tubular epithelium is so reduced that absorption and secretion are markedly impaired. In other words, at low temperatures the excised kidney is acting almost entirely as a filter.

The glomerular membrane does not normally allow large amounts of plasma protein to pass, although small quantities of protein are present in glomerular filtrate. Since protein cannot be detected in the urine in health it must be concluded that such protein as passes through the glomerular membrane is reabsorbed by the tubules, and there is evidence that this is so. The presence of protein in the urine, a common finding in renal disease, implies an increased leakage through the glomeruli rather than defective tubular reabsorption. However, proteinuria is not infrequently found where there is no renal disease. Increased pressure in the renal veins probably accounts for the proteinuria of pregnancy and also the proteinuria found in about 5 per cent. of healthy young adults. This 'orthostatic albuminuria' occurs only when the subject is erect, disappearing when he lies down. The stagnant anoxia of cardiac failure and the anæmic anoxia of chronic anæmia are associated with the passage of quite large amounts of protein in the urine, but whether this is due solely to the anoxia has not been clearly established. The proteinuria so commonly detected in fever, chronic infective processes and malignant disease likewise awaits a clear explanation.

About 125 ml. of fluid are filtered through the glomerular membranes of the two kidneys per minute (p. 604). Since the surface area of the glomerular capillaries is about 1 sq. metre the amount filtered per sq. cm per minute is only $125 \div 10^4 = 0{\cdot}0125$ ml. Since there are about 2×10^6 glomeruli, about 0·06 cu.mm must be filtered through each per minute.

GLOMERULAR PRESSURE.—The pressure in the afferent arteriole of the glomerulus of the frog has been measured directly by Hayman. A micropipette connected to a manometer was first inserted into the fluid space in a glomerular capsule and, after blocking the neck of the tubule by pressing on it with a glass rod (Fig. 34, 6), the pressure in the capsule was raised first until (1) the flow of blood was arrested in some of the capillaries of the glomerulus and secondly until (2) no red cells entered the glomerulus from the afferent arteriole. Measurement (1) is taken as the mean pressure in the glomerulus, usually about 50 per cent. of aortic pressure, and (2) is the afferent arteriolar pressure, usually 85 per cent. of aortic pressure (which is 21 to 61 cm water in the frog). The pressure in the glomerulus is thus adequate to account for the production of a protein-free filtrate of the plasma. Indirect measurements by Winton of glomerular pressure in the excised dog kidney give a value of 70 per cent. of the arterial pressure (see also p. 597).

Effects of Circulatory Changes on Glomerular Function.—If the fluid in the glomerular capsule is really an ultrafiltrate of plasma then a rise in arterial pressure might be expected to augment the volume of filtrate and an increase in the ureteric pressure, a back pressure, to diminish it. While both these expectations have been shown to be correct, experiments designed to verify them may be complicated in several ways. In the frog, for example, microscopic examination shows that many glomeruli become temporarily inactive by reason of marked constriction of the glomerular arterioles. Later the constriction passes off and filtration is resumed, while at the same time adjacent glomeruli may become inactive. This intermittency of glomerular function, implying the existence of a reserve of renal units, probably does not occur in mammalian kidneys which are none the less capable of varying within very wide limits the volume of urine excreted. Even if there is never complete constriction of the glomerular arterioles in the mammalian kidney the pressure in the glomerular capillaries can be increased or decreased by alteration in the relative degrees of vasoconstriction of the afferent and efferent glomerular arterioles, with the results indicated in Table 34, 7. Extensive studies of the relationship between renal blood flow and glomerular filtration rate have not yet definitely established the roles of the afferent and efferent vessels in the total renal resistance, but in man the musculature of the efferent arteriole is probably the main regulator of blood flow through the kidney.

TABLE 34, 7

Effect of Constriction of Glomerular Arterioles on Glomerular Function

Constriction of the afferent arteriole produces:	Constriction of the efferent arteriole produces:
Reduction in renal blood flow. Fall of glomerular pressure. Reduced glomerular filtration. *Little change in filtration fraction.	Reduction in renal blood flow. Rise of glomerular pressure. Increased glomerular filtration. *Increased filtration fraction.

Changes in glomerular filtration may involve secondary tubular effects.
*The term 'filtration fraction' is explained on p. 607.

If there is no vasoconstriction in the afferent glomerular arterioles the effect of raising the systemic blood pressure is to increase the amount of the filtrate and so give less time for any modification of its composition as the filtrate passes down the tubules (*pressure diuresis*). An isolated kidney perfused with blood first at a normal and then at a raised arterial pressure produces more urine under the latter conditions, and this urine contains more chloride and less creatinine than when the pressure is lower. The increased output of urine is accompanied by an increase in oxygen consumption, an indication that the tubule epithelium is performing more work. If the greater reabsorption by the tubules does not keep pace with the augmented glomerular filtration rate the urinary output rises. The increased volume of urine produced when the arterial pressure is raised is thus not the result solely of the operation of simple physical principles.

When a rabbit kidney is perfused at a constant volume per minute the addition of adrenaline to the perfusion fluid causes a rise of perfusion pressure from constriction of the efferent arterioles. At the same time there is an increase in kidney volume, the result of the accumulation of blood in the glomerular tufts. The increase of pressure in the glomerulus augments filtration and may lead to an increased production of urine. However, no matter what the volume of the glomerular filtrate, the actual volume of the urine is governed almost entirely by the amount of water reabsorbed by the tubules. Even when glomerular filtration is markedly reduced the volume of filtrate still considerably exceeds the amount of urine produced in a given time, so that the output of urine is not a reliable indication of renal function.

Effective Filtration Pressure.—The hydrostatic pressure of the blood in the glomerular capillaries forces water and solutes through the glomerulus against the osmotic pressure of the plasma proteins. Pressure is also necessary to force the filtrate from the glomerular capsule along the lumen of the tubule. Thus the effective filtration pressure P_{ef} is given by

$$P_{ef} = P_g - P_{pp} - P_c$$

where P_g = blood pressure in glomerular capillaries,
P_{pp} = osmotic pressure of plasma proteins (about 25 mm Hg),
and P_c = intrarenal pressure, that is the smallest pressure which must be applied to the ureter in order to reduce urine flow (about 20 mm Hg).

From this it can be seen that P_{ef} can be reduced to zero, that is filtration ceases, by lowering the arterial pressure (and therefore P_g) or by raising ureteric pressure, and thus increasing P_c. Since P_{ef} is then zero, P_g must equal $P_{pp} + P_c$. In other words, P_g, the pressure in the glomerular capillaries, must always exceed 45 mm Hg. The level to which the arterial pressure must be lowered to suppress urinary flow varies according to the conditions of the experiment, but even the lowest values, about 40 mm Hg, are in excess of the osmotic pressure of the plasma proteins. By comparing the effects of raising venous and ureteric pressures Winton calculated that the pressure in the glomerular capillaries in the excised mammalian kidney is about 70 per cent. of the aortic pressure. This is more than enough to overcome the osmotic pressure of the plasma proteins. There is thus no doubt that in normal circumstances the head of pressure in the glomerulus is amply sufficient to overcome the colloid osmotic pressure. If, however, the blood pressure falls considerably the output of urine may cease.

The equation above makes it clear that P_{ef} rises when P_{pp} falls. A reduction in the colloid osmotic pressure of the plasma can be brought about by the rapid intravenous injection of normal saline; the resultant increase in the effective filtration pressure produces a large increase in urinary output, much larger in fact than can be explained on the basis of filtration alone. It is likely that the increased output is due to a change in tubular activity.

TUBULAR FUNCTION

The filtration of fluid at the glomerulus is, then, an essential step in the formation of urine since the tubular cells cannot function if there is no glomerular filtrate. The function of the tubular epithelium is to some extent determined by the amount and composition of the filtrate because tubular activity depends, among other things, upon the concentration of solutes in the fluid within them, upon its osmotic pressure and upon its rate of flow along the tubules. Glomerular filtration is governed by the blood flow through, and the pressure within, the glomerular capillaries, and depends also on the structural integrity of the renal corpuscle. Gross alterations in one or all of these factors lead to defective filtration which may, however, be concealed by decreased tubular reabsorption or active tubular secretion.

Samples of glomerular filtrate obtained by micropipettes are, as we have seen, filtrates of blood plasma but, since the concentrations of substances in these filtrates are very different from their concentrations in urine, the composition of the tubular fluid must be altered considerably during its passage down to the collecting tubules. Having accepted the fact of glomerular filtration we must now attempt to account for these alterations, the extent of which is shown in Table 34, 11 *b*.

Tubular Reabsorption.—In the frog kidney Richards found that the amount of fluid obtainable by pipette from each glomerular capsule is about 1 cu. mm per hour. The 2000 capsules of each frog kidney produce, therefore, about 2 ml. of filtrate per hour, although only 0·2 ml. of urine appears in that time. Thus some 90 per cent. of the water of the glomerular filtrate does not leave the kidney but is reabsorbed by the tubules. This cannot be the only tubular activity since some substances are present in about the same concentration in blood and urine whereas others, such as urea and creatinine, exist in much higher concentration in the urine. These differences could be easily explained if both absorption and secretion occurred in the tubules.

The use of micropipettes (Fig. 34, 8) has made possible the examination of the fluid within the tubules of the frog kidney at various points in its passage along the lumen. In the proximal tubule, where there is little alteration in the concentration of salts and glucose in the glomerular filtrate, water and electrolytes are reabsorbed as an iso-osmotic solution, while in the distal tubule water can be reabsorbed against an osmotic gradient and the osmotic pressure of the urine can rise considerably above that of the plasma. To the latter process, the mechanism of which is quite unknown, the name 'facultative reabsorption' of water has been given, to distinguish it from the 'obligatory reabsorption' which occurs in the proximal tubule. In the distal tubules also the concentration of hydrogen ions in the tubular fluid is adjusted to the needs of the body (p. 642). Since the concentration of urea in the urine of the frog is higher than can be produced solely by reduction in volume of the glomerular filtrate it must be concluded that in this animal additional urea is actively transferred (secreted) by the cells of the tubules into the tubular lumen.

Similar experiments with micropipettes have been performed on guinea-pigs and rats, but only the proximal tubules of the mammalian kidney can be reached. It has been found that the proximal tubule absorbs amino-acids and glucose but not creatinine; it also absorbs sodium and therefore chloride and water. Urea filtered at the glomeruli diffuses from the tubules back into the plasma. This is a passive process depending on the fact that the concentration of urea in the tubular fluid exceeds its concentration in the plasma. The rate of 'back-diffusion' of urea is inversely related to the urine flow; in other words the elimination of urea from the body is least when the urine is most concentrated and the flow along the tubules the slowest.

The capacity of the tubular epithelium to reabsorb the various constituents presented to it in the glomerular filtrate (*the maximum reabsorptive capacity or* Tm) is limited and has a different value for different materials. Normally the whole of the glucose filtered by the glomeruli is reabsorbed by the tubules; if, however, the plasma concentration of glucose is gradually raised by an intravenous infusion then a concentration of glucose in the plasma is eventually reached, about 180 mg. per 100 ml. in a healthy adult, at which the cells of the tubules are presented with glucose faster than they can absorb it. Glucose then begins to be eliminated in the urine. This limiting concentration of glucose in the blood used to be referred to as the 'renal threshold' for glucose. It will be realized, however, that it represents a critical rate of delivery of glucose to the tubules, that is to say, it depends not only upon the plasma concentration but also on the glomerular filtration rate. It follows, therefore, that when the latter is much reduced, as for example in a severely dehydrated diabetic patient, the urine may contain no glucose even though the concentration of glucose in the blood is much above the normal. A method for measuring the glucose Tm is given on p. 605.

The reabsorption maxima of Na^+, K^+, Ca^{++}, Mg^{++} and Cl^- are less than for glucose, and they are consequently normally present in urine.

A second group of substances present in glomerular filtrate, uric acid and phosphates for example, passes back through the tubular epithelium like urea by a simple process of diffusion when their concentration in the lumen of the tubules exceeds that in the plasma. Such substances do not normally appear in lower concentration in the urine than in the plasma.

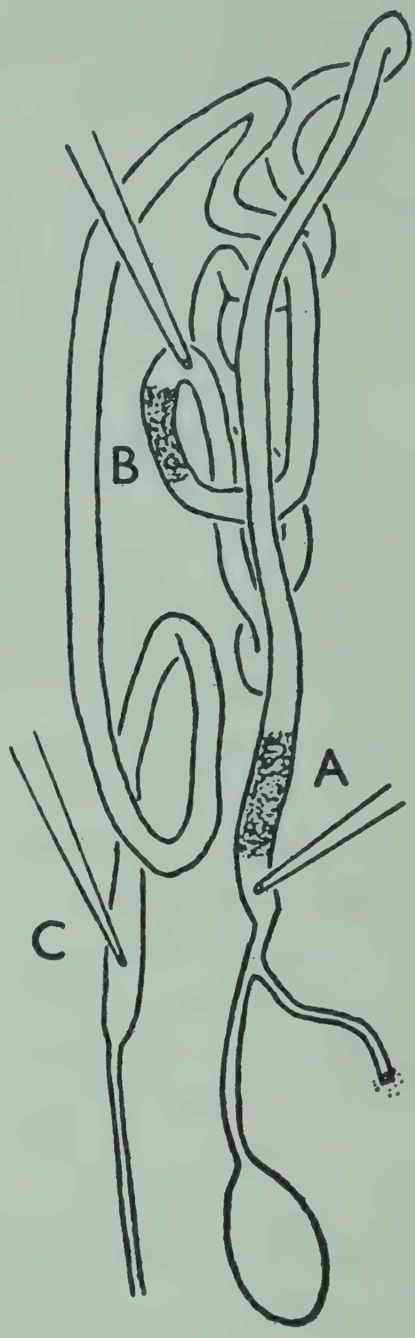


Fig. 34, 8

The insertion of a micropipette for the collection of tubular fluid from the beginning A, the middle B and the end C of a first convoluted tubule of Necturus. Stippling indicates mercury inserted to block the lumen (A. N. Richards & A. M. Walker (1937), *Amer. J. Physiol.* **118,** 114).

Finally creatinine, unique in this respect among substances normally found in the body, is not reabsorbed by the tubules. In man, though not in the dog, there is active secretion of creatinine by the tubule epithelium. In other words, in man creatinine is eliminated both at the glomerulus by filtration and by the tubules by secretion.

Tubular Secretion.—Although there is little doubt that reabsorption is the main function of the renal tubules, there is also evidence that secretion occurs. In those fishes in which the nephrons have no glomeruli (aglomerular kidneys) the urine, which contains the usual urinary constituents such as chlorides, creatinine and uric acid, must be formed entirely by secretion. The aglomerular kidney can also secrete phenol red injected into the blood stream, but not protein or glucose. In the dog's kidney phenol red is excreted at a greater rate than can be accounted for by simple reabsorption and in man the organic iodine compound, diodone (diodrast), is certainly excreted by the tubules, for the reason that more of this substance appears in the urine in a given time than is filtered through the glomerular membrane.

Although there is abundant evidence for the actual secretion by the tubule cells from the blood into the lumen of the tubules of many 'foreign' substances, such as phenol red and diodone, the only physiological materials that may appear in the urine in greater amounts than can be accounted for on the basis of concentration of the glomerular filtrate are creatinine, potassium and hydrogen ion. It is therefore believed that these are secreted by the tubules. The relation between the secretion of hydrogen and potassium ions is in fact so close that it seems likely that they compete for the same component of the transport mechanism.

The renal tubules thus have a considerable range of functional activity. From glomerular filtrate by selective reabsorption and excretion they elaborate urine the amount, composition and reaction of which are adjusted to the needs of the body. By synthesis of ammonium ion and its exchange with kation in the glomerular filtrate and by excretion of hydrogen ion they play an important part in acid-base regulation (p. 642).

In concentrating the glomerular filtrate the cells of the tubules must do a certain amount of osmotic work, and since their capacities in this respect are limited there is a maximal concentration which can be achieved or, to put it another way, there is a minimum amount of water which must be excreted for a given amount of solids. Since the molecules remaining in the tubules must be accompanied by water to dissolve them, substances such as urea, sodium sulphate and sodium phosphate, if given in excess, act as *diuretics*, that is, as substances which increase the volume of urine. In the same way substances like glucose, which are in normal concentrations easily absorbed, cannot be completely absorbed if present in great excess and diuresis results. This accounts for the diuresis which occurs in diabetes mellitus when the blood sugar is high. The phenomenon is referred to as *osmotic diuresis*.

WORK OF THE KIDNEY

Since the glomerular membrane is a simple filter, no work is done by the kidney in the filtration process, the energy for which is supplied by the heart. The transfer of materials from the tubules to the plasma is done against an osmotic gradient and requires the expenditure of an amount of energy which can be calculated for each urinary constituent. The total osmotic work done is about 0·7 kcal per litre of urine, or 1 kcal per day.

The oxygen usage is estimated at 3 ml./g./hour or (for two human kidneys each weighing 150 g.) 22 litres/300 g./day. Since 1 litre of oxygen is equivalent to about 5 kcal (p. 154) the efficiency of the kidney is $\frac{1}{22 \times 5} = \frac{1}{110}$, or less than

1 per cent. Even when due allowance is made for the errors involved in this rough calculation the mechanical efficiency is much lower than that calculated for other organs. We must remember, however, that although the kidney, regarded simply as an organ for producing urine, may be considered to have a low efficiency, it has a number of other useful functions, such as the elimination of toxic materials, the regulation of acid-base balance and the osmotic pressure of the blood. The work involved in these processes cannot yet be determined but they are none the less essential to the smooth working of the body.

THE VOLUME AND COMPOSITION OF THE URINE

Even in health the volume and composition of the urine may vary widely from day to day, such variations being governed by, for example, the type of food consumed, the volume of fluid taken and the amount of fluid loss by other channels—a factor which itself depends on environmental temperature and humidity, exercise and sweating (p. 203). These physiological variations make it difficult to give any exact figures for the normal range of the chemical composition of the urine and even to make any absolute distinction between physiological and pathological constituents, since a substance which may normally be present in small amounts in urine may, when excreted in large amounts, assume pathological importance.

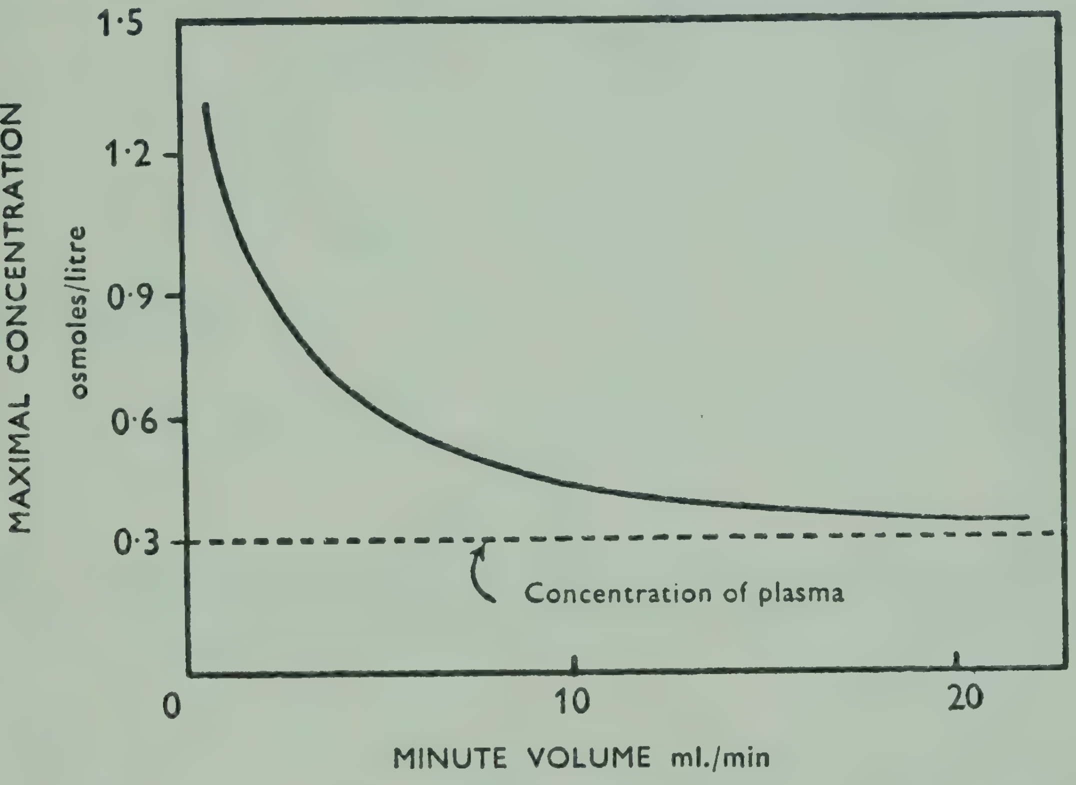


Fig. 34, 9

Relation of maximal urinary concentration to minute volume of urine (J. R. Robinson (1954), *Reflections on Renal Function.* Oxford: Blackwell Scientific Publications.

The yellow colour of normal urine is due mainly to the pigment urochrome. The chemical nature and source of this substance are unknown but it probably arises endogenously, since its output increases with tissue protein destruction or when metabolism is increased. Traces of other pigments, such as urobilin and uroerythrin, are also normally present. The colour of the urine in general varies according to its concentration. Freshly passed urine possesses a characteristic aromatic odour but when it is allowed to stand it develops an ammoniacal odour, owing to the bacterial decomposition of urea to ammonia.

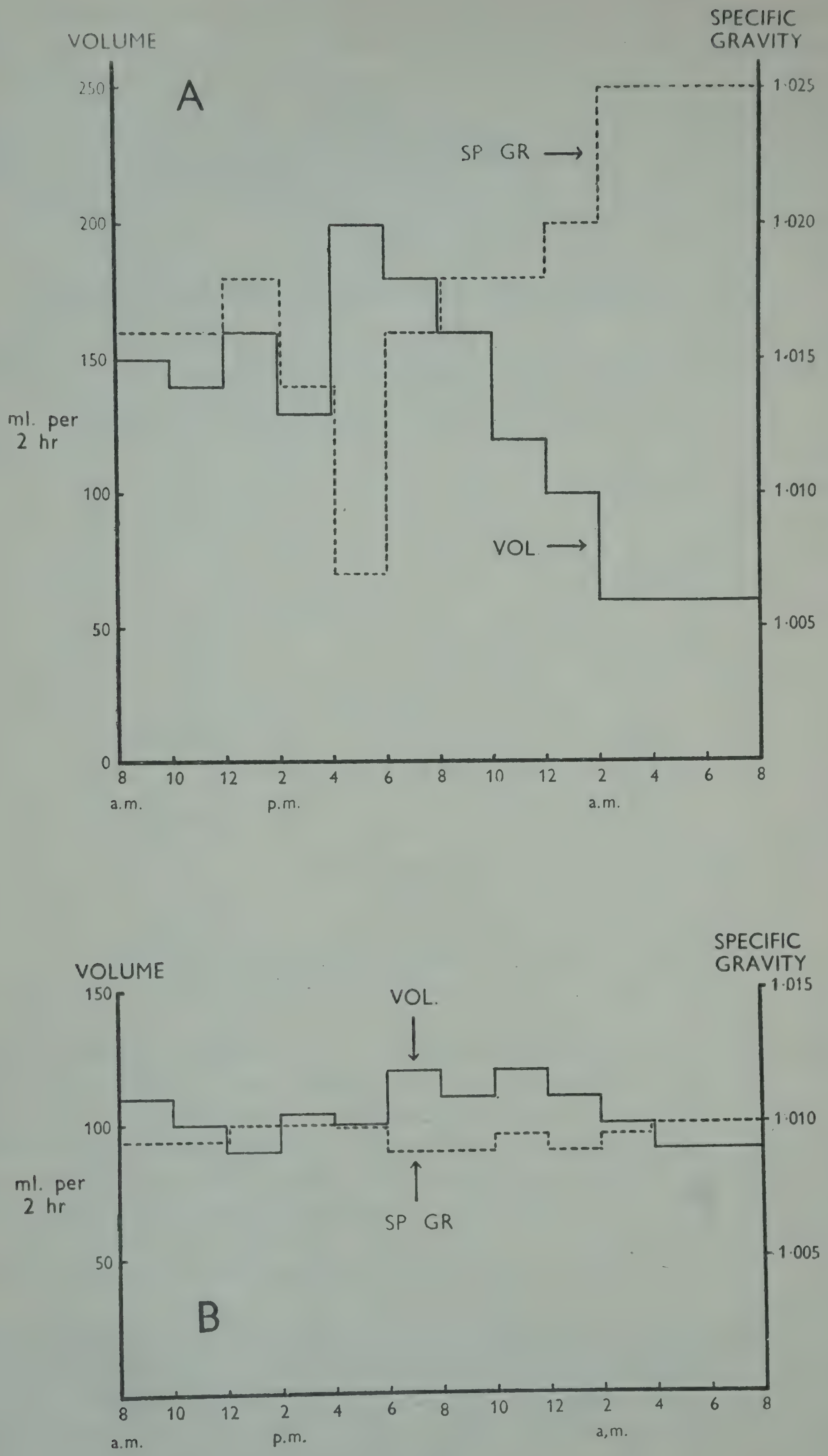


FIG. 34, 10

Diurnal variations in volume of urine and specific gravity in a healthy person (A) and in a patient with advanced renal disease (B). The volume of urine produced during the night is relatively small in A; there is no such change in B.

The specific gravity of the urine in health can vary between 1·001 and 1·040 according to the concentration of dissolved solids, although it usually lies between 1·010 and 1·025 representing the elimination of some 40 to 60 g. of dissolved solids in 24 hours. Interpreted in terms of pressure these figures mean a variation from 0·3 to 1·4 osmoles per litre (Fig. 34, 9) according to the urine flow. In renal disease the kidney becomes unable to alter the concentration of the urine to meet the needs of the body and the specific gravity remains fixed about 1·010 (*isosthenuria*); this represents an osmolarity of about 0·3 osmoles per l., the same as that of plasma.

TABLE 34, 11 *a*

Principal Constituents of Urine. Grams per 24 hours

Urea	20 to 30	≡ 9 to 14 g. nitrogen
Uric acid	0·6	≡ 0·2 g. nitrogen
Creatinine	1·2	≡ 0·4 g. nitrogen
Ammonia	0·5 to 0·9	≡ 0·4 to 0·7 g. nitrogen
Sodium	6	
Potassium	2	
Calcium	0·2	
Chloride, as Cl	7	
Phosphate, as P . . .	1·7	
Sulphate, as H_2SO_4 . . .	1·8	

These figures are approximate only. The range of variation in healthy persons is considerable.

TABLE 34, 11 *b*

To show the Difference in Constitution between a Day's Urine and the Glomerular Filtrate from which it was derived

Substance	Amount in Glomerular Filtrate g.	Amount Eliminated in the Urine g.
Sodium	600	6
Potassium	35	2
Calcium	5	0·2
Glucose	200	trace
Urea	60	35
Water	180 litres	1·5 litres

J. R. Robinson (1954), *Reflections on Renal Function.* Oxford: Blackwell Scientific Publications.

Normally about half of the total daily fluid intake is eliminated through the kidneys, accounting for a daily output of urine between 1000 and 1600 ml. The urine is not, however, produced at a constant rate during the 24 hours, for the volume of urine during the day (8 a.m. to 8 p.m.) is two to four times greater than that secreted during the night (8 p.m. to 8 a.m.) (Fig. 34, 10, A). An increase in the night urine relative to the day urine is often an early indication of renal disease because when the kidney is damaged the hourly flow of urine cannot be varied.

The average concentration of the principal urinary constituents has already been referred to in Table 34, 5, but because the main factor governing alterations in concentration is the amount of water excreted it is often of more value to consider the quantity of each constituent eliminated in 24 hours (as given in Table 34, 11 *a*).

CLEARANCE TESTS

In the investigation of renal function in man it is important to be able to estimate the amount of deproteinized plasma which passes into the glomerular capsule, that is the glomerular filtrate. This is possible if a non-toxic substance can be found which is small enough to pass through the glomerular membrane but yet is neither excreted into the tubules nor absorbed from the tubular urine. Inulin, a polysaccharide of molecular weight 5200, containing some 30 fructose residues, meets these requirements.

The information required for the calculation is the volume of urine V passed in ml. per minute, the concentration P of inulin in the plasma in g. per 100 ml. and its concentration U in the urine in g. per 100 ml. The amount of inulin excreted per minute is obviously $\frac{UV}{100}$ g. Since the glomerular fluid is an ultrafiltrate of blood plasma each ml. of glomerular filtrate contains $\frac{P}{100}$ g. inulin. If x is the volume of glomerular filtrate per minute then x ml. of glomerular filtrate contain $\frac{xP}{100}$ g. inulin. But $\frac{UV}{100}$ g. of inulin have been excreted per minute, therefore $\frac{xP}{100}=\frac{UV}{100}$ or $x=\frac{UV}{P}$ ml. per min.

An arithmetical example may make the idea of clearance more readily appreciated. Suppose the plasma to contain 50 mg. (or 0·05 g.) of inulin per 100 ml. and the urine 6·25 g. inulin per 100 ml., while the urine is excreted at the rate of 1 ml. per min. The inulin clearance or glomerular filtration rate is then

$$\frac{6{\cdot}25 \times 1}{0{\cdot}05} = 125 \text{ ml./min.}$$

This value is referred to as the *plasma clearance* for inulin, since it can be thought of as that volume of plasma from which all the inulin has been removed (cleared) in one minute. The term plasma clearance is a general one that can be applied to any substance in the plasma which is excreted in the urine, no matter whether the actual process is filtration, secretion, or excretion.

The evidence that the inulin clearance actually measures the glomerular rate is as follows. The plasma clearance of inulin is not affected by the concentration of inulin in the plasma. The concentration of inulin in the glomerular filtrate obtained by direct puncture in the frog is the same as its concentration in the plasma, and the large molecule of inulin is not likely to diffuse out of the renal tubules. Aglomerular kidneys do not secrete inulin. The plasma clearance of inulin is the same as that for sorbitol and mannitol, which have much smaller molecules (molecular weight 182 in both cases). Furthermore, the inulin clearance in the dog is equal to the clearance of creatinine, ferrocyanide and thiosulphate. These facts are difficult to explain unless all these substances are unabsorbed on their way down the tubules. If an animal or a man is injected with the glycoside phloridzin, glucose appears in the urine, since the tubules are then unable to reabsorb glucose. The glucose and inulin clearances are then for practical purposes identical; this suggests that in these circumstances both are filtered and no tubular reabsorption occurs. When the blood pressure is too low for glomerular filtration inulin does not appear in the tubular fluid, but phenol red which is secreted by the tubules does appear.

Now the amount of water in ml. reabsorbed by the tubules per minute is $\frac{UV}{P} - V$, that is, the volume of the glomerular filtrate less the actual volume of urine produced in the same time. The amount of water reabsorbed from each ml. of glomerular filtrate is

$$\frac{\frac{UV}{P} - V}{\frac{UV}{P}} = 1 - \frac{P}{U}$$

Thus if the concentration of inulin is known in samples of blood and urine taken at the same time the power of the kidney to reabsorb water can be calculated. Using the values just given ($U = 6{\cdot}25$ g./100 ml. and $P = 0{\cdot}05$ g./100 ml.) the amount of water reabsorbed from each ml. of glomerular filtrate is

$$1 - \frac{0{\cdot}05}{6{\cdot}25} = 1 - \frac{1}{125} = 0{\cdot}992 \text{ ml.}$$

Although in some animals creatinine can be used to estimate the volume of glomerular filtrate, in man this substance is excreted in part through the tubules, a fact which accounts for the finding that in man the creatinine clearance is some 30 to 40 per cent. higher than the clearance of inulin.

The maximum rate of tubular reabsorption of glucose (glucose-Tm) in man is measured by raising the plasma glucose by intravenous injection to a high level and then measuring the inulin clearance and the glucose excretion. If the inulin clearance is 125 ml. per minute when the plasma sugar is 400 mg. per 100 ml. then $125 \times \frac{400}{100} = 500$ mg. of glucose is filtered per min at the glomeruli, but 150 mg. of glucose is found in the urine per minute. The maximum tubular reabsorption of glucose (glucose-Tm) is, therefore, $500 - 150 = 350$ mg. per minute.

The substance diodone (diodrast), 3:5-diiodo-4-pyridon-N-acetic acid, is secreted so completely by the tubules that after an intravenous injection very little is found in the renal veins. The plasma clearance of diodone calculated by the method just given for inulin is, therefore, much higher than that of inulin, and is very nearly equal to the total plasma flow through the kidneys. If the plasma contains 5 mg. (0·005 g.) of diodone iodine per 100 ml. and the urine contains 3·5 g. of diodone iodine per 100 ml., while the volume of urine is 1 ml. per minute, these values can be substituted in the formula $\frac{UV}{P}$. The diodone clearance is, therefore, $\frac{3{\cdot}5 \times 1}{0{\cdot}005} = 700$ ml. per minute. If the inulin clearance measured at the same time is 125 ml./min then $\frac{125 \times 5}{100} = 6{\cdot}25$ mg. of diodone iodine are filtered off at the glomerulus and the remainder $\frac{(700-125)5}{100} = \frac{575 \times 5}{100} = 28{\cdot}75$ mg. is excreted by the tubules. Direct comparison in dogs and in man of the composition of the blood from the renal artery and renal vein shows that the kidney extracts from the plasma 85 to 90 per cent. of the diodone in the renal artery. This method, therefore, somewhat underestimates the true renal plasma flow and it is strictly applicable only when the extraction is known to be relatively complete. The total blood flow through the kidneys can be calculated if the percentage of plasma and corpuscles is measured in a hæmatocrit; that is the blood flow equals the renal plasma flow divided by the percentage of plasma and multiplied by 100. Just as there is a maximal reabsorptive capacity for each substance reabsorbed from the glomerular filtrate, so there is a maximum secretory capacity, that is to say, the rate at which materials can be transferred from the blood across the tubular epithelium into the lumen is limited. If the plasma level of diodone (or creatinine, or phenol red) exceeds a certain level, which is different for each substance, the tubular capacity to secrete it is completely taken up, so that any additional increase in the concentration of diodone in the plasma does not lead to increased secretion. The concentration of diodone in the urine increases, however, because each ml. filtered at the glomerulus contains more diodone.

If the diodone is expressed as iodine its Tm is equivalent to 57 mg. iodine per minute. Because the secretory powers of the tubules are limited diodone clearances do not estimate renal flow unless the plasma concentration is below 10 mg./100 ml. If the renal plasma flow is 700 ml./min when the glomerular filtration rate is 125 ml./min 575 ml. of plasma pass through the capillaries supplying the tubules. If this contains 10 mg. per 100 ml. the tubules will just be able to clear out the 57·5 mg.

The sodium salt of *p*-amino-hippuric acid may also be used to estimate plasma flow because, like diodone, it is almost completely cleared from the plasma.

Estimation of Active Kidney Tissue.—Since glucose is first filtered and then reabsorbed, estimation of glucose-Tm gives a clue to the number of normally functioning nephrons in the kidneys. Glucose-Tm is reduced if the glomeruli are damaged and cease to filter, or if the tubules are damaged and cease to reabsorb.

Since diodone excretion is carried out chiefly by the tubules, diodone-Tm is little affected by damage to the glomeruli and diodone-Tm is, therefore, referred to as the *tubular excretory capacity*.

Homer Smith has measured glucose-Tm and diodone-Tm in man both during the vasoconstriction produced by adrenaline and after the injection of a foreign protein which produces fever and vasodilatation.

The two values did not differ greatly from the normal. This and other evidence strongly suggests that the intermittent glomerular activity characteristic of the frog kidney does not occur in man.

The ratio $\frac{\text{inulin clearance}}{\text{diodone clearance}}$, that is, $\frac{\text{glomerular filtrate}}{\text{plasma flow through kidney}}$, is called the *filtration fraction.* It amounts to some 20 per cent. of the renal plasma flow and represents the proportion of the latter filtered per minute.

The following are approximate figures for a normal man:

	ml./min
Blood flow through the kidneys	1200
Plasma flow through the kidneys (diodone clearance)	700
Glomerular filtrate (inulin clearance)	125
Urine	1
Filtration fraction $125 \div 700 \times 100$	18 per cent.

The value for urea clearance, about 75 ml. per minute, is much lower than that for inulin, since some of the urea diffuses back into the plasma. The urea clearance measured during diuresis produced by drinking a pint of water or a solution of urea is used as an index of kidney function. If the flow of urine is small the urea clearance is reduced because when the flow along the tubules is slow more urea passes back into the blood. Urea clearance is greatly reduced in nephritis; in gross kidney disease it may be as low as 3 or 4 ml. per minute.

The conception of plasma clearance is illustrated diagrammatically in the accompanying diagram, Fig. 34, 12. The range of plasma clearance is from zero for glucose, at normal plasma levels completely absorbed, up to 700 ml./min for diodone, which is almost entirely removed from the plasma.

In animals renal blood flow can be measured by collecting the blood from the renal vein, or more elegantly by shunting this blood for a short period into a volume recorder. The Fick principle (p. 449) can also be applied to the kidney just as to the heart. It is necessary to know the plasma concentration of a substance in the arteries C_A and in the veins C_V and the concentration in the urine C_u. The amount of the substance lost in the urine in one minute is $\frac{V \times C_u}{100}$.

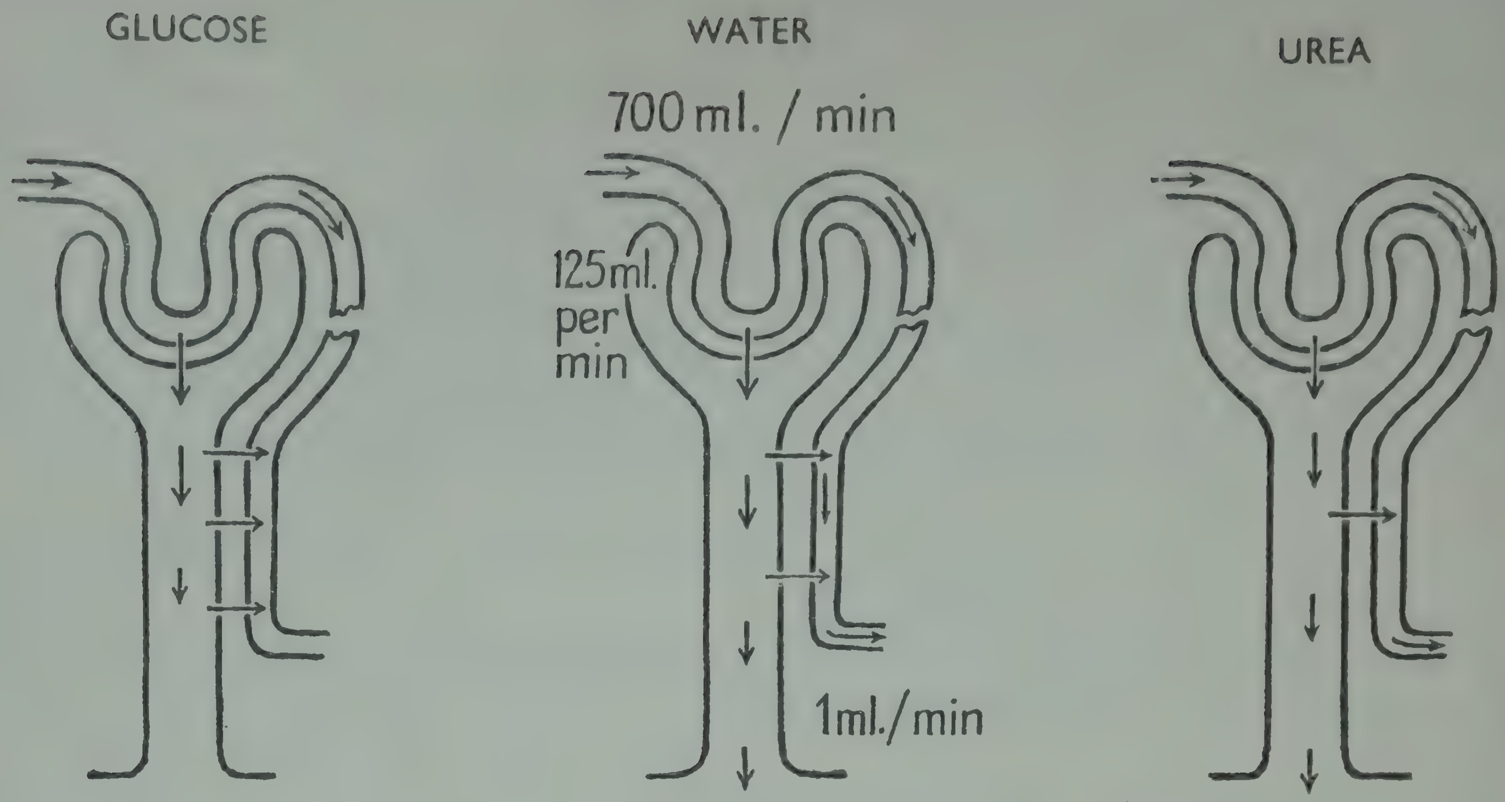
Each 100 ml. of plasma loses $C_A - C_V$ g.

$\frac{100}{C_A - C_V}$ ml. of plasma loses 1 g.

$\frac{100}{C_A - C_V} \times \frac{V \times C_u}{100}$ ml. of plasma loses $\frac{V \times C_u}{100}$ g.

The plasma flow is, therefore, $\frac{V \times C_u}{C_A - C_V}$ ml. per minute.

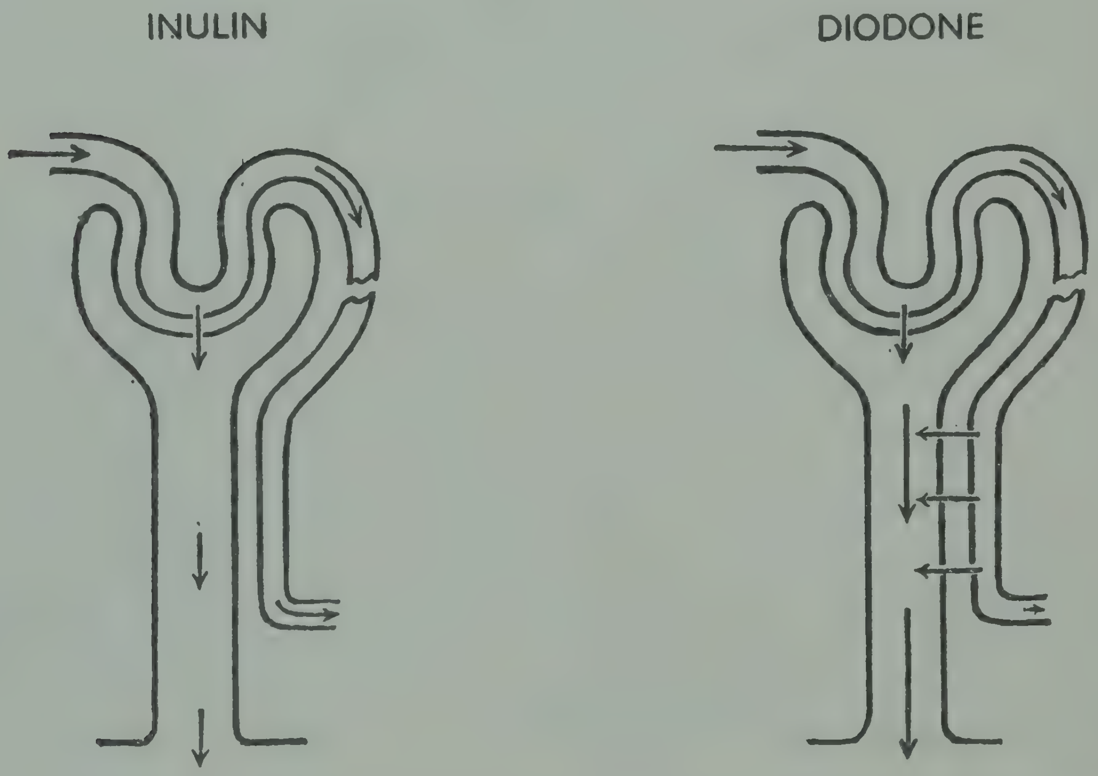
If all of a substance in the plasma of the renal arteries is eliminated by the kidneys and none reaches the veins $C_V = 0$ and the plasma flow becomes $\frac{V \times C_u}{C_A}$. If for C_u we put U and for C_A we put P we arrive at the formula $\frac{UV}{P}$ which was used before for calculating clearance. The Fick principle,



Glucose is completely reabsorbed. Plasma clearance = 0.

Water is reabsorbed very nearly completely (99 per cent.). Plasma clearance = 1 ml./min.

Urea is partially reabsorbed. Plasma clearance = 75 ml./min.



Inulin is neither reabsorbed nor excreted by the tubules. Plasma clearance is glomerular filtrate = 125 ml./min.

Diodone is filtered at the glomerulus and excreted into the tubules so that only a small amount leaves the kidney in the renal veins. Plasma clearance nearly equals plasma flow = 700 ml./min.

Fig. 34, 12

Diagram of a nephron with afferent and efferent glomerular vessels and a tubular capillary. The arrows show the direction of movement of the substance under discussion and their length indicates the amount which is moving. The middle diagram above shows the amount of fluid passing into the kidney (700 ml. per min), the amount filtered at the glomerulus (125 ml. per min) and the amount eliminated as urine (1 ml. per min). Since the 700 ml. of plasma flowing through the glomerulus per min carries the red cells the volume of blood flowing through the kidney is 1200 ml./min.

though more suitable for animals, can be applied to man if a cardiac catheter is passed through the right atrium along the inferior vena cava and into one renal vein.

The clearance tests devised by Homer Smith and his group have made it possible to investigate renal physiology in unanæsthetized normal man. The subject drinks water on the evening before the experiment and in the morning so that his tissues are well hydrated and high urine flows are obtained. Inulin and diodone are given by continuous intravenous infusion. Fig. 34, 13, shows an experiment recorded by Homer Smith in which a large amount of adrenaline was given. The renal plasma flow (diodone clearance D) was reduced from 926 ml. per min to a minimum of 510 ml. per min, although the mean blood pressure (BP) was probably unaffected. Obviously vasoconstriction had taken place and the site can be inferred thus. Narrowing of the afferent arterioles (see Table 34, 7) would have produced a fall of glomerular pressure with little change in the filtration fraction. In fact the filtration fraction (FF) rose and the filtration (IN) was almost unaffected. This can be explained only by a constriction of the efferent glomerular arteriole. Experiments of this kind show that glomerular filtration rate in man is remarkably constant under basal conditions. Considerable variations can however occur in

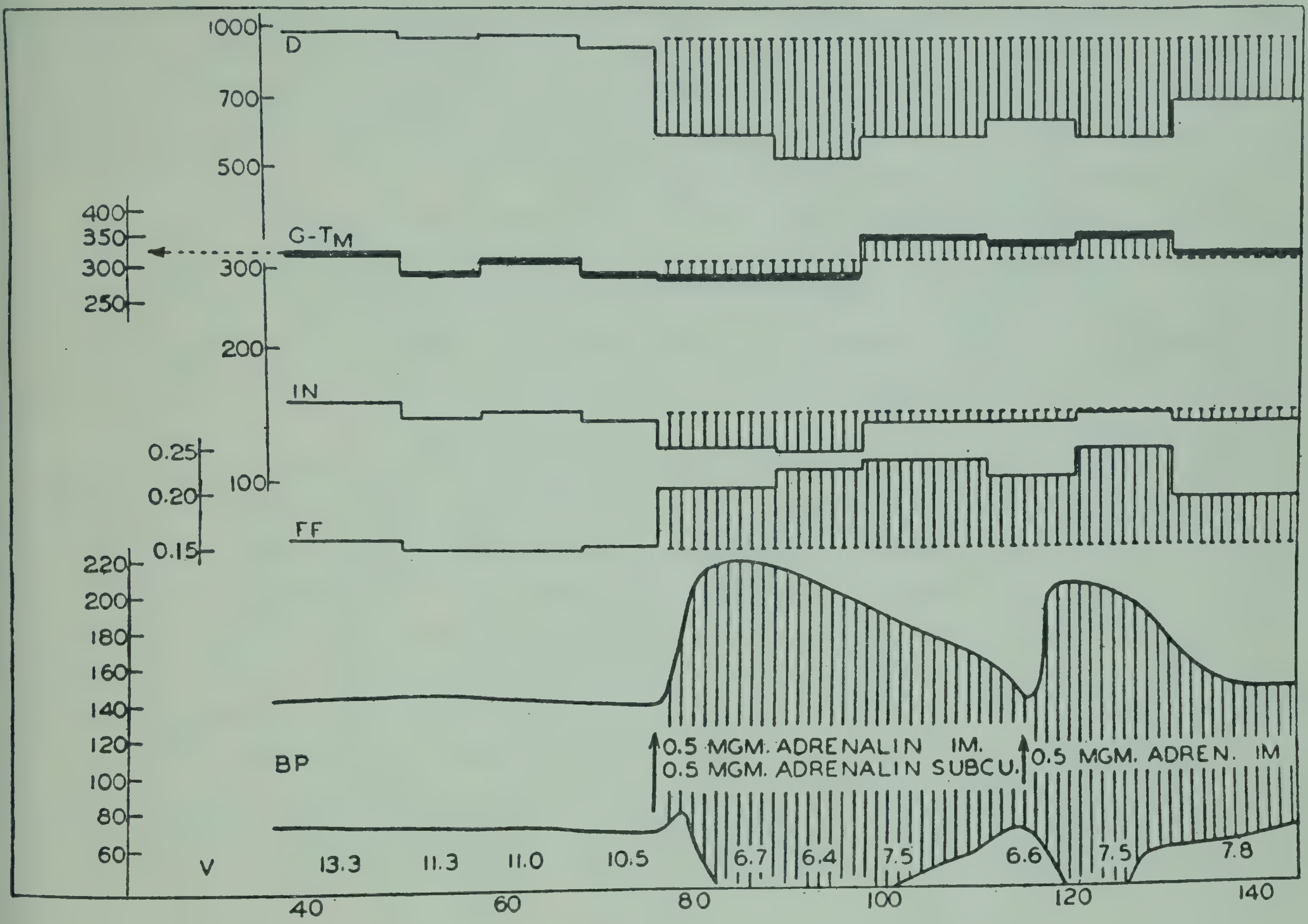


FIG. 34, 13

Action of adrenaline on effective renal plasma flow, etc. D = diodone clearance in ml. per minute. IN = inulin clearance in ml. per minute. Filtration fraction, FF = inulin/diodone clearance ratio. Blood pressure BP in mm Hg and urine flow V in ml. per minute. $G-T_M$ = maximal rate of tubular reabsorption of glucose under conditions of tubular saturation, in mg. per minute. $G-T_M$ is indirect index of the number of active glomeruli in the kidney (Homer W. Smith (1943), *Lectures on the Kidney*. Lawrence, Kansas: University of Kansas).

response to alteration of arterial pressure and during exercise, as well as in dehydration and with disease. Variations in urinary output are largely dependent, however, on variations in tubular reabsorption of water. The glucose-Tm (G–Tm) was measured during the experiment illustrated in Fig. 34, 13, and was almost unaffected (306 mg. per min before adrenaline and 325 mg. per min after). This shows that no glomeruli were shut down by this large dose of adrenaline.

Constriction of the efferent arterioles with a marked reduction in urinary output has been found to occur when the subject of the experiment becomes alarmed. This reaction may be due to the outpouring of adrenaline from the suprarenal medulla or to impulses passing along the sympathetic nerves to the kidneys. The relation between emotion and urinary output is discussed also on p. 619.

Renal vasoconstriction also occurs in syncope produced by standing still in the upright position, but here the constriction affects the afferent arterioles. In traumatic shock there is also renal vasoconstriction; since the kidneys normally receive a very large part (about one-quarter) of the cardiac output this vasoconstriction may be beneficial by diverting blood to other organs.

Dilatation of the renal vessels with increase of the rate of flow through the kidney is not easy to produce. Cutting off vasoconstrictor impulses by high spinal anæsthesia does not produce renal hyperæmia. There are thus normally no tonic vasomotor impulses to the kidney. The only method so far discovered of achieving a large increase in diodone clearance is to give a dose of a foreign protein which produces a rise of body temperature (artificial pyrexia). During the hyperæmia the filtration fraction is reduced, showing that the efferent glomerular arterioles are dilated. This reduction offsets the increase in blood flow so that the filtration rate (inulin clearance) is not much altered. This hyperæmia occurs also in denervated kidneys but the reason for it is not yet known.

RELATION OF THE KIDNEYS TO BLOOD PRESSURE

A persistent elevation of mean arterial pressure (high blood pressure, p. 486) could be accounted for by a rise in cardiac output, by an increase in blood volume or blood viscosity, or by a generalized increase of peripheral arterial resistance. It is now agreed that in this condition the first three of these factors remain within normal limits; the persistently elevated arterial pressure can be accounted for only by an increase of peripheral vascular resistance. As this is not due to actual narrowing (stenosis) of the blood vessels, it must be the result of an increased peripheral resistance.

Disease of the kidneys in man is often accompanied by high blood pressure and in 1933 Goldblatt succeeded in producing a sustained increase of blood pressure by experiments on animals. If a clamp is applied to one renal artery in a dog so as to reduce the blood flow considerably (*ischæmia*), the blood pressure rises slowly for a month or so. The increased blood pressure depends on the presence of the ischæmic kidney because removal of the latter is followed by a fall of pressure to the normal level (Fig. 34, 14).

Surgical removal of one grossly diseased kidney has very occasionally been successful in relieving high blood pressure in man.

Many experiments support the idea that the elevated arterial pressure which follows renal ischæmia is due to the liberation into the blood of a constrictor substance. Thus, for example, clamping of the renal arteries produces hypertension even after division of all the renal and splanchnic nerves. Clamping of the renal artery of a kidney transplanted into the neck and receiving its blood from the animal's carotid artery is followed by an increase in blood pressure. The kidney of a donor dog is first made ischæmic by a renal clamp so that a rise of blood pressure is produced; some time later another dog is nephrectomized and then has the ischæmic kidney connected to its circulation by joining its carotid artery to the renal artery and its jugular vein to the renal vein. In the course of quarter of an hour the blood pressure of the recipient rises and remains high until the graft is removed, when it slowly declines. If a normal kidney is connected in the same way to a nephrectomized recipient there is seldom any rise of pressure.

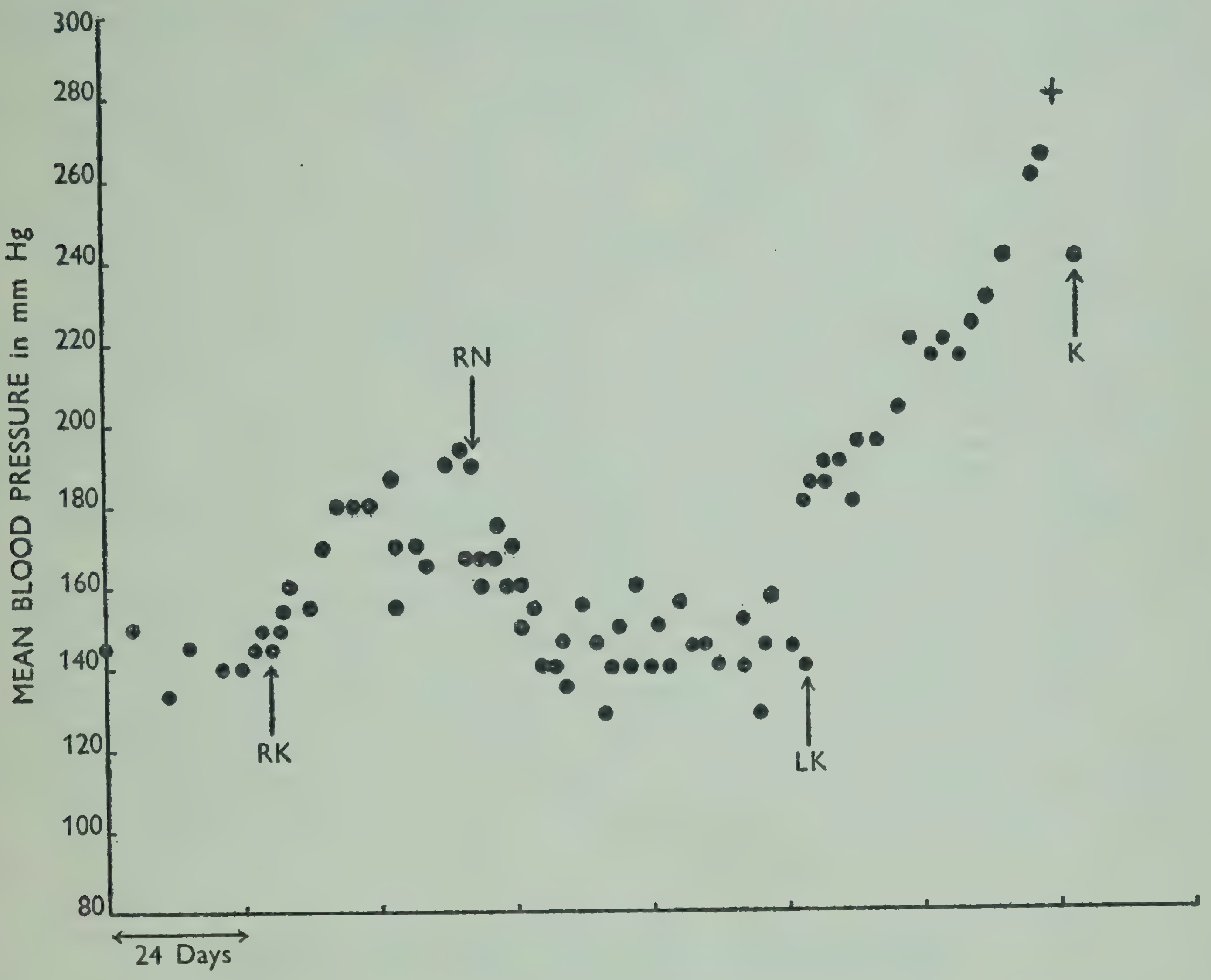


FIG. 34, 14

Experimental hypertension in a female dog of 17·8 kg. RK = right main renal artery moderately constricted. RN = right nephrectomy at a time when the mean blood pressure was elevated resulted in its prompt return to normal. LK = severe constriction of main renal artery of left kidney. This was followed by very high elevation of mean blood pressure. + = the mean blood pressure at this time was more than 300 mm Hg. K = killed (H. Goldblatt (1937–38), *Ann. intern. Med.* **11,** 75).

Chemical Mechanism of Renal Ischæmic Hypertension.—In 1898 Tigerstedt and Bergman observed a prolonged rise of blood pressure in the unanæsthetized rabbit after intravenous injection of extracts of normal renal cortex. To the active principle in these extracts they give the name *renin*. Renin is not itself a pressor substance; perfusion of the arteries of the isolated rabbit ear with Locke's solution containing purified renin does not produce a vasoconstriction. If serum as well as renin is added to the artificial perfusion fluid then the arteries contract. The action of renin is due to the fact that it is a proteolytic enzyme capable of hydrolysing a normal constituent of the α-globulin fraction of the plasma protein (*hypertensinogen*) with the production of a new vasoconstrictor, and therefore pressor, body called *hypertensin*.

Hypertensin is a heat-stable polypeptide with a molecular weight of about 2700; it has been prepared in a relatively pure form and gives a crystalline picrate and oxalate. The injection of hypertensin into an animal is followed by an immediate increase in blood pressure which returns to the basal level, in the rabbit, within 10 to 15 minutes.

Renin has been detected in blood withdrawn from the renal vein of animals at a time when their blood pressure was low as the result of hæmorrhage or of the infusion of vasodilator substances; it is possible, therefore, that the kidneys have a role in the homoiostatic regulation of blood pressure. It seems clear, however, that high blood pressure in man is not produced through the renin-hypertensin mechanism.

The significance of this hypertensive mechanism in normal physiology is not at present clear. Renal ischæmia may merely unmask a mechanism which is otherwise difficult to detect. Whether the healthy kidneys have an important role in the homoiostatic regulation of blood pressure through the humoral mechanism of renin production has not yet been established, although evidence is accumulating that such may be the case.

REGULATION OF WATER BALANCE

Since 120 ml. of glomerular filtrate are produced each minute and the total volume of extracellular fluid in the body is about 15 litres, the whole of the fluid outside the body cells passes through the kidneys every 15 minutes. By variations in glomerular filtration rate and in tubular reabsorption and secretion the kidney is able to exercise a great influence on the water and electrolyte composition of the blood plasma, and normally these variations keep its chemical composition and reaction remarkably constant. These regulatory functions could be carried out either if the kidney itself possessed the ability to react to small changes in the composition of the blood circulating through it, or if renal function were controlled by distant chemoreceptors or osmoreceptors (p. 617) through hormonal or nervous influences. It is only in the case of the regulation of osmotic pressure that we have much knowledge of the actual mechanism.

Response to Ingestion of Water and Salt.—An increase in the amount of urine follows increased intake of water (Fig. 34, 15). After it has been absorbed the water reduces the tonicity of the extracellular fluid, and there follows a copious secretion of dilute urine, the maximal rate of elimination occurring twenty or thirty minutes later than the maximal dilution of the plasma. Because the reabsorption of solids in the tubules is only slightly decreased the specific gravity of the urine is much reduced. Both the inulin and the diodone clearances remain unaltered during the diuresis which must therefore be the result of diminished tubular reabsorption of water.

When isotonic saline is drunk the diuresis is insignificant (Fig. 34, 15), the extra fluid being eliminated only slowly; in this case there is no change in the tonicity of the extracellular fluid. Large infusions of isotonic saline given rapidly do, however, produce a diuresis, the result of the sudden expansion of the extracellular space.

Effect of the Posterior Pituitary Gland.—An excised perfused kidney produces only very dilute urine but, if an extract of the posterior lobe of the pituitary gland is added to the perfusion fluid, the urine becomes more concentrated, its volume diminishes and the concentration of chloride in it increases. If, instead of adding pituitary extract, the blood perfusing the excised kidney is first passed through a head with an intact pituitary gland the same fall in the volume of urine and increase in sodium chloride concentration is observed (Fig. 34, 16). Patients with the disease called *diabetes insipidus* pass large quantities, up to thirty litres or more per day, of very dilute urine; in other words, their renal function approaches that of the isolated perfused kidney. In such patients the injection of posterior lobe extracts restores the output of urine to normal. Diabetes insipidus is associated with pathological lesions at the base of the brain and in the pituitary body, and the syndrome can be produced experimentally in animals by injury to the posterior lobe of the pituitary or to the tracts which connect it with the hypothalamus (supra-optico-hypophysial tracts) (Fig. 34, 17).

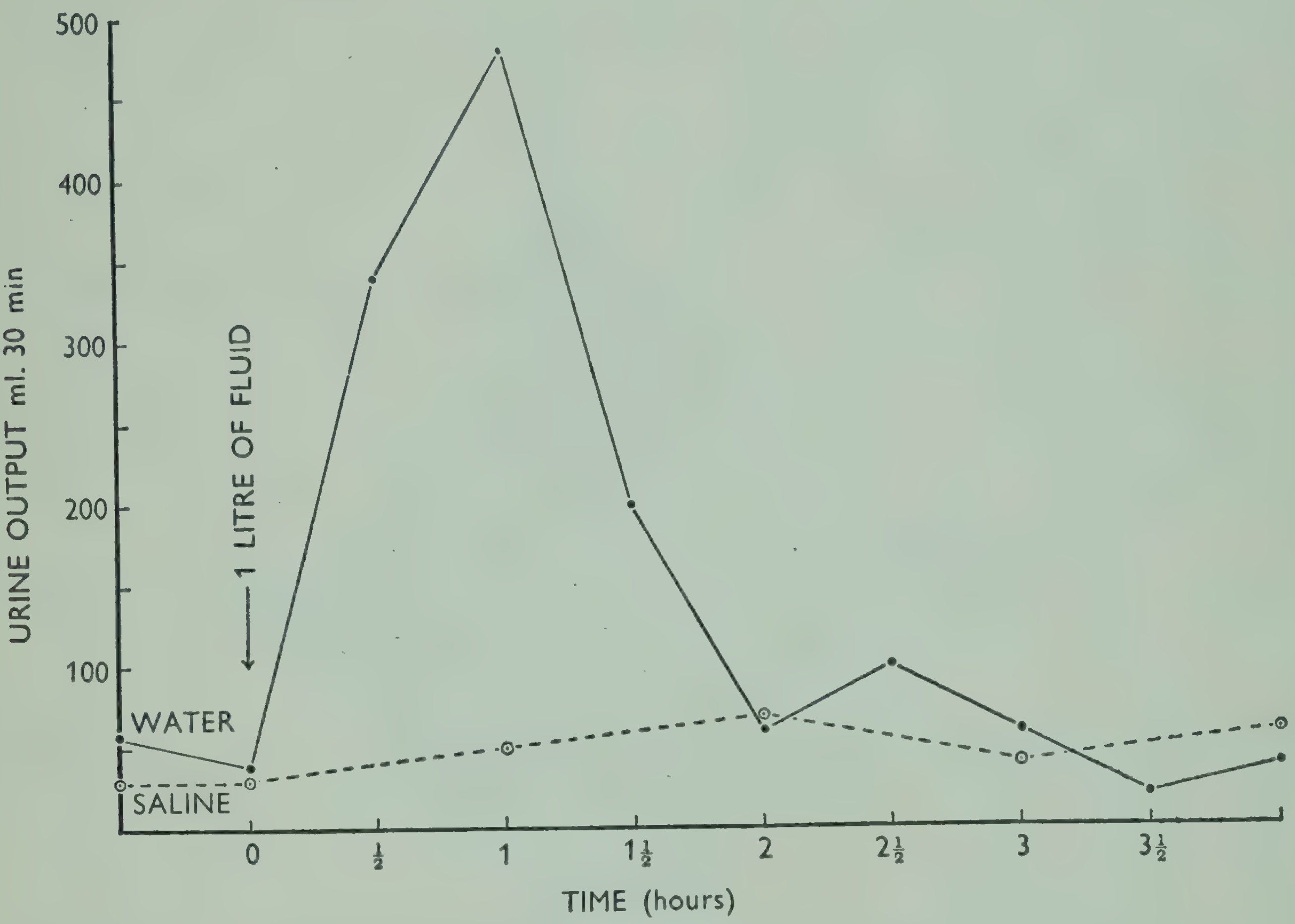


Fig. 34, 15

Response to the drinking of 1 litre of water and, on the following day, 1 litre of 0·9 per cent. sodium chloride solution by a healthy man. At the end of two hours all the water, but little of the saline, had been eliminated.

The hypothesis that the posterior pituitary produces an *antidiuretic hormone* (ADH) which circulates in the blood to govern renal function can explain these facts; normally a relatively small volume of concentrated

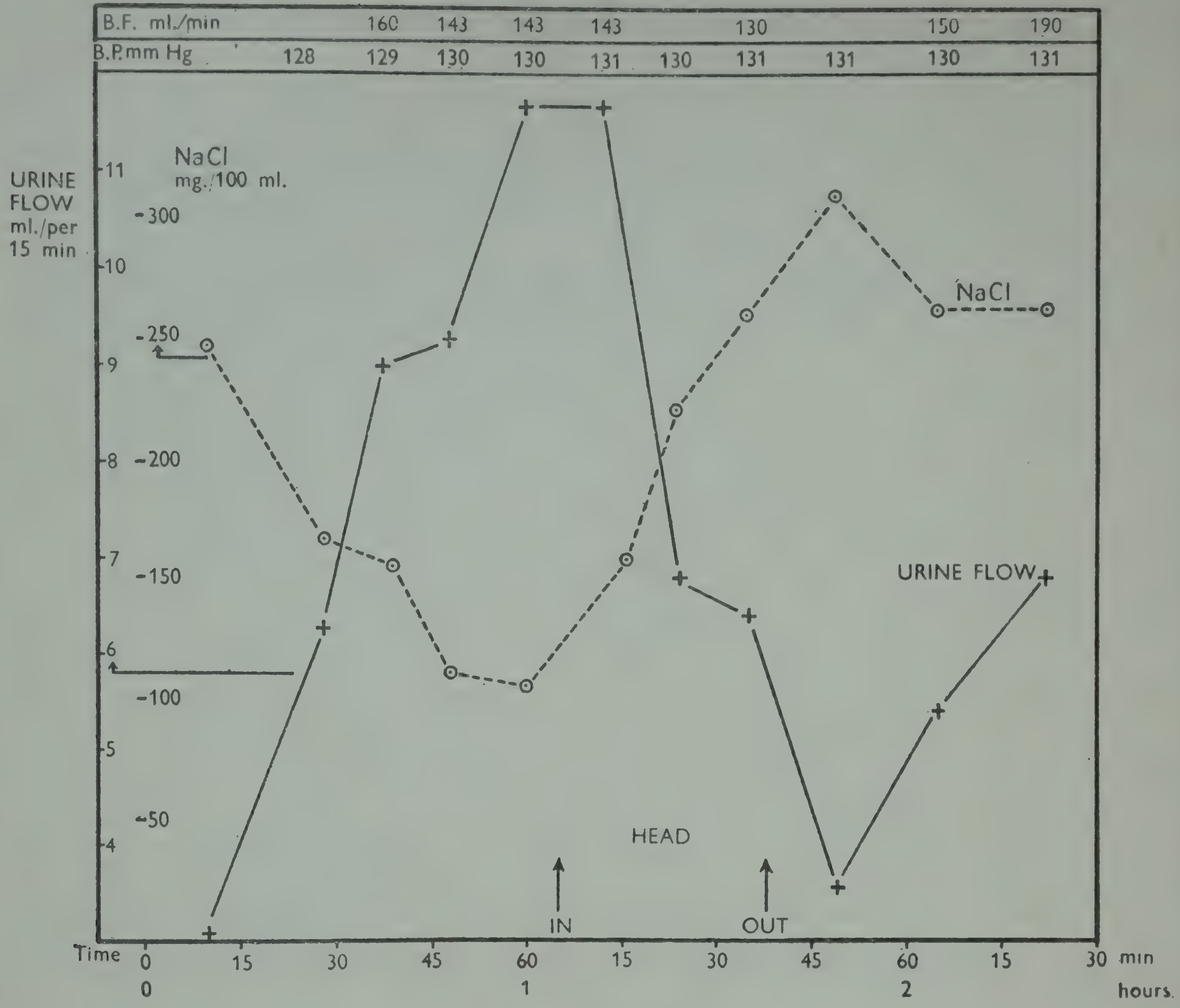


Fig. 34, 16

Urine flow in ml. per 15 minutes and sodium chloride concentration in mg. per 100 ml. of urine; dog's kidney in heart-lung-kidney perfusion preparation. A head was switched into the perfusion circuit in parallel with the kidney and out at the arrows so marked. Above, renal blood flow (BF) in ml. per minute and blood pressure in mm. Hg (E. B. Verney (1926), *Proc. Roy. Soc.* B, **99,** 497).

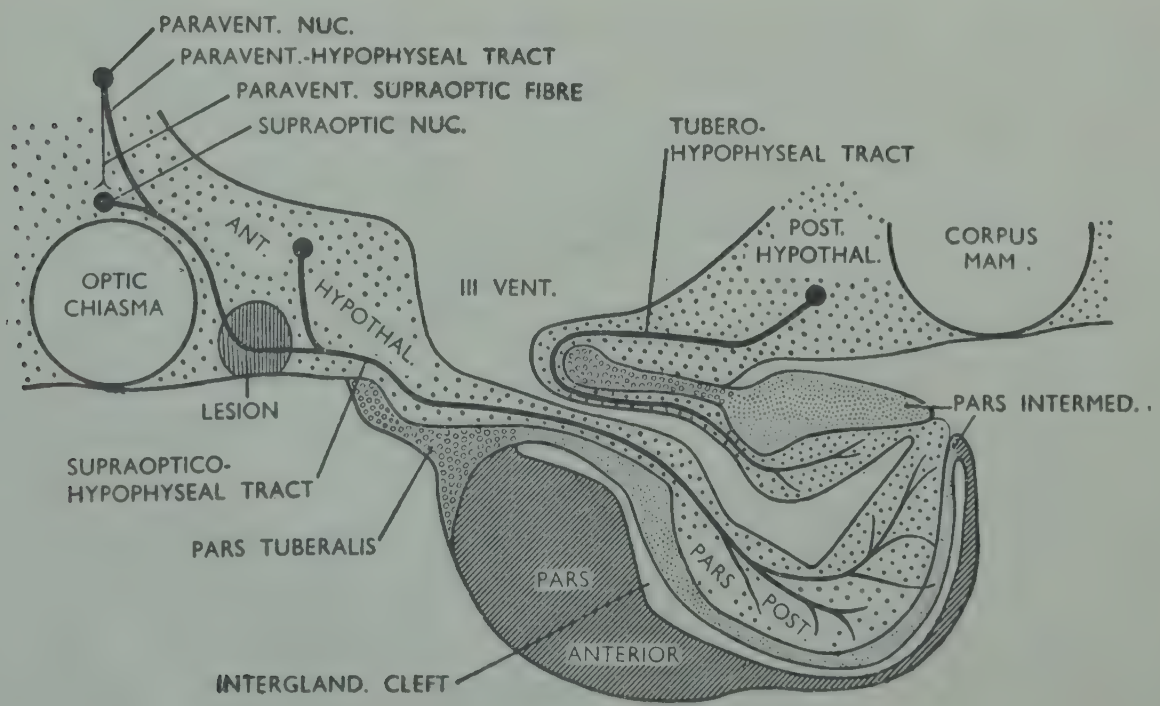

PARAVENT. NUC.
PARAVENT.-HYPOPHYSEAL TRACT
PARAVENT. SUPRAOPTIC FIBRE
SUPRAOPTIC NUC.
TUBERO-HYPOPHYSEAL TRACT
POST. HYPOTHAL.
CORPUS MAM.
OPTIC CHIASMA
ANT. HYPOTHAL.
III VENT.
LESION
PARS INTERMED.
SUPRAOPTICO-HYPOPHYSEAL TRACT
PARS TUBERALIS
PARS ANTERIOR
PARS POST.
INTERGLAND. CLEFT

Fig. 34, 17

The nerve connexions of the pituitary gland of the cat. The lesion necessary for the production of diabetes insipidus is indicated (C. Fisher, W. R. Ingram, W. K. Hare & S. W. Ranson (1935), *Anat. Rec.* **63,** 31).

urine is produced but in the absence of the hormone the urine is copious and dilute. A dog with experimental diabetes insipidus has less power of concentrating the glomerular filtrate than a normal dog and the urine, because of the greatly reduced absorption of water in the distal convoluted tubule, is hypotonic. When water is withheld the hypophysectomized animal can, however, produce a hypertonic urine. This suggests that although the antidiuretic hormone enhances the power of the tubular cells to absorb water against an osmotic gradient it is not essential for the process. The chief action of the antidiuretic hormone is to increase water reabsorption in the distal convoluted tubules. There is an associated rise in urinary chloride concentration but the total amount of chloride eliminated is not altered.

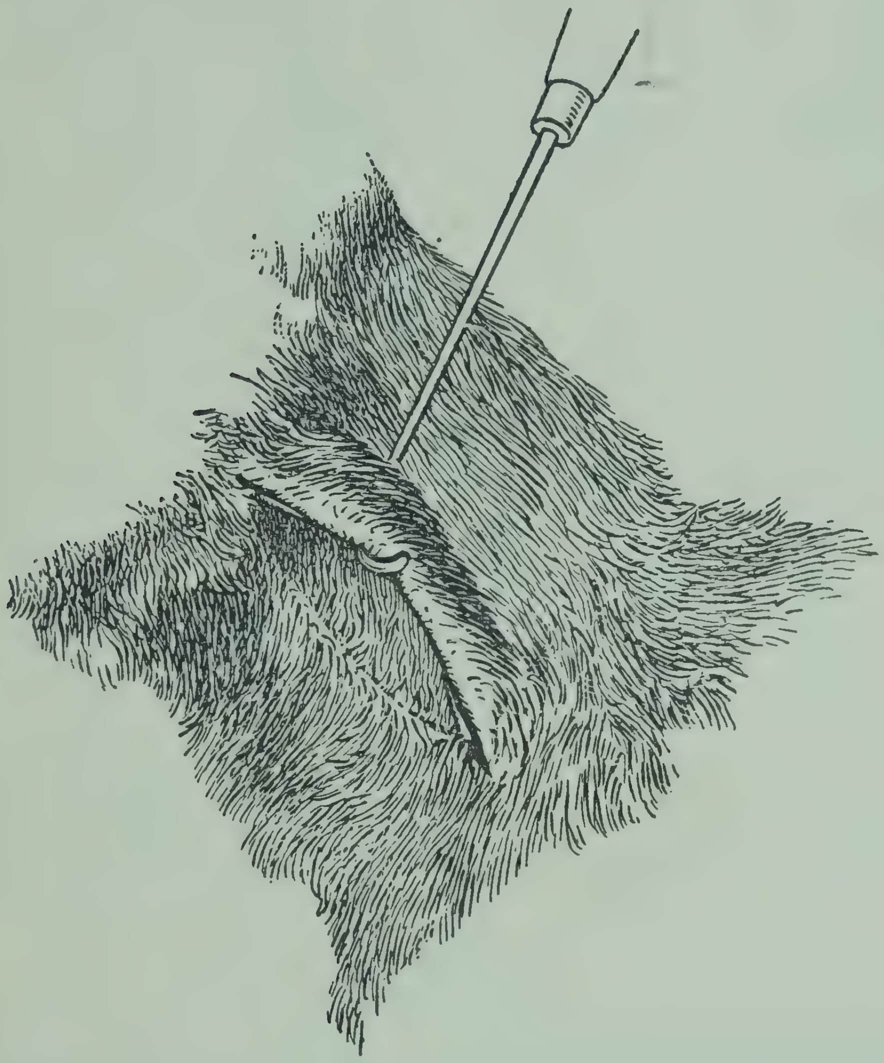


FIG. 34, 18

Carotid loop in the neck of a dog.

A clue to the actual amount of antidiuretic hormone secreted under normal circumstances is given by the finding that a dog (10 to 15 kg.) deprived of its posterior pituitary lobe but given an intravenous infusion of 1 to 5 milliunits of pituitary extract per hour reacts to ingestion of water like a normal animal.

Effect of Hypertonic Solutions.—Much information about the secretion of the antidiuretic hormone has recently been obtained by Verney. He dissected out the common carotid artery of his dogs and enclosed it in a tube of skin to form a 'carotid loop' (Fig. 34, 18). It was then easy to inject drugs or solutions directly into the carotid artery. The urine was collected by catheters into graduated flasks. Control injections of isotonic solutions of various substances either intra-arterially or intravenously had no effect on urinary output, but hypertonic solutions of sodium chloride, glucose, sucrose and sodium sulphate all produced the same fall of urinary output (Fig. 34, 19, A) provided that they were given in a standard short time, say ten seconds. The effect could be imitated by injection of small quantities of posterior extract and the amount of this required to match the original response could be found by trying various doses, as shown in Fig. 34, 19, D, E and F.

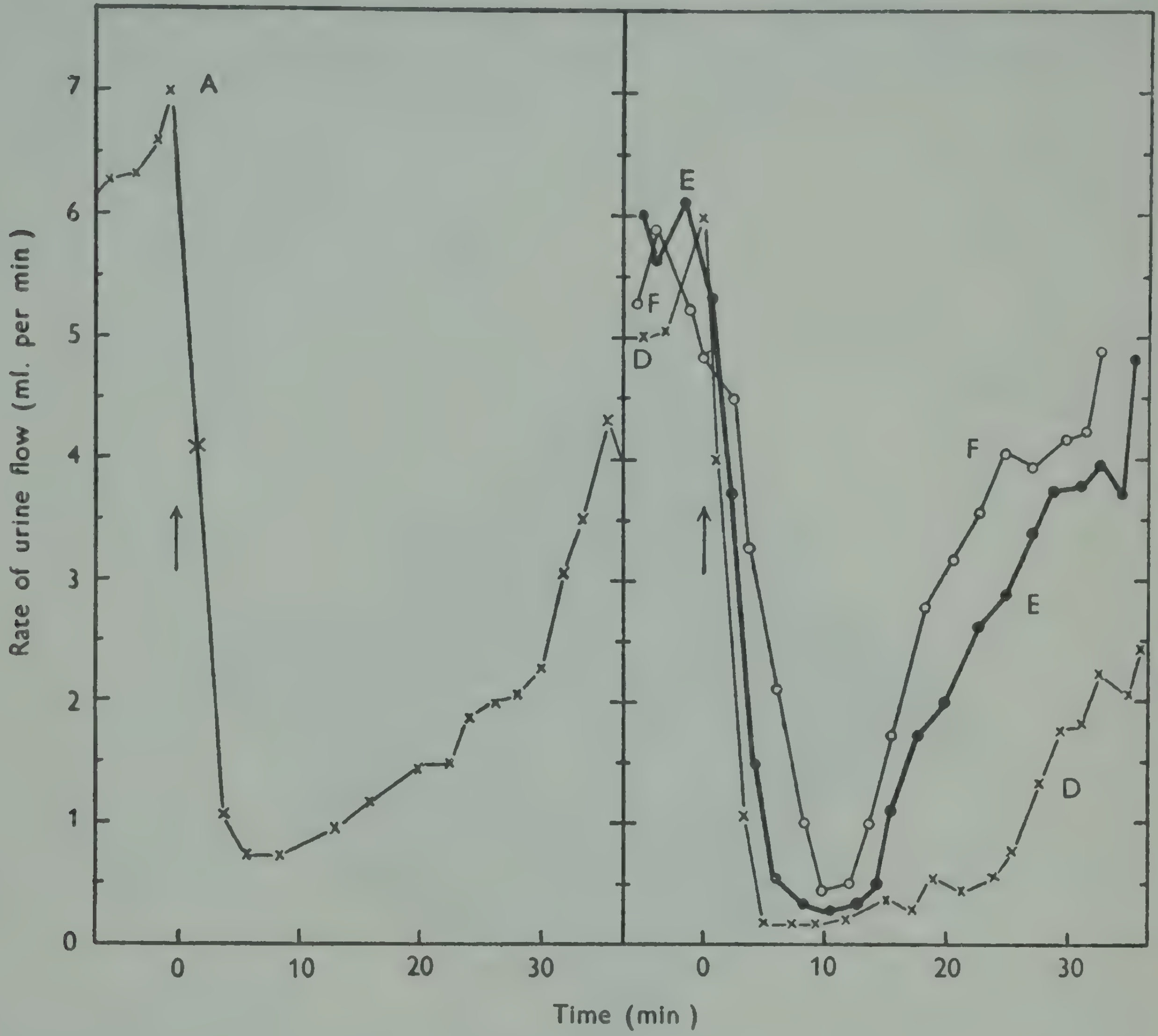


FIG. 34, 19

The inhibition of water diuresis by a rise of osmotic pressure in the carotid artery. At A 10 ml. of 2·5 per cent. sodium chloride was injected into the right carotid artery in 12 seconds. Injection of 10 ml. of 15·4 per cent. glucose in 11 seconds had almost the same effect. On the right an attempt is made to imitate the antidiuretic effect of the hypertonic solution by injecting at the arrow, D, 3 mU; E, 2 mU, and F, 1 mU of posterior lobe extract into the malleolar vein (1 mU = 1 milliunit) (E. B. Verney (1947), *Proc. Roy. Soc.* B, **135**, 67).

The response to the injection of such hypertonic solutions is considerably diminished, though not abolished, by removal of the pituitary body. In Fig. 34, 20, the effect of injecting hypertonic sodium chloride into the carotid artery of the intact animal (*a*) is compared with that produced by the same injection after removal of the posterior lobe of the pituitary (*c*). The antidiuretic response does not disappear entirely, but to judge from the dose of pituitary extract required to match the effect of the hypertonic solution (*b* and *d*) it is reduced to one-tenth. The small residual response is probably due to the activity of tissue around the stalk of the pituitary gland in the hypothalamic area of the brain. Although there is an increase in urinary output for a short time after hypophysectomy the remaining secretory tissue is able in a few days to keep the urinary output at the pre-operative level. In other words, hypophysectomy, since it does not remove all the tissue producing the antidiuretic substance, results only in a temporary diabetes insipidus.

The slow injection of hypertonic sodium chloride over a period of half an hour or so instead of a few seconds gave results that were essentially similar. When hypertonic salt solution was injected into the carotid artery the antidiuretic effect appeared sooner and was greater in amount than when the injection was made into a vein in a hind-leg. During an intravenous injection of hypertonic sodium chloride the concentration of salt in the blood is quickly reduced by admixture with venous blood arriving from all parts of the body at the right side of the heart and, therefore, the rise in osmotic pressure in the carotid arteries is correspondingly smaller than that which follows injection of hypertonic solution directly into the latter vessel. This strongly suggests that the receptors sensitive to alterations in osmotic pressure (*osmoreceptors*) lie somewhere in the peripheral supply of the carotid arteries. Since ligation of the internal carotid branch destroys all antidiuretic response to the injection of hypertonic solution it is clear that the osmoreceptors must lie in its distribution, and it is in fact probable that they are concentrated in the supraoptic nucleus which can control the secretion of the posterior lobe through the supraoptico-hypophyseal tract (Fig. 34, 17).

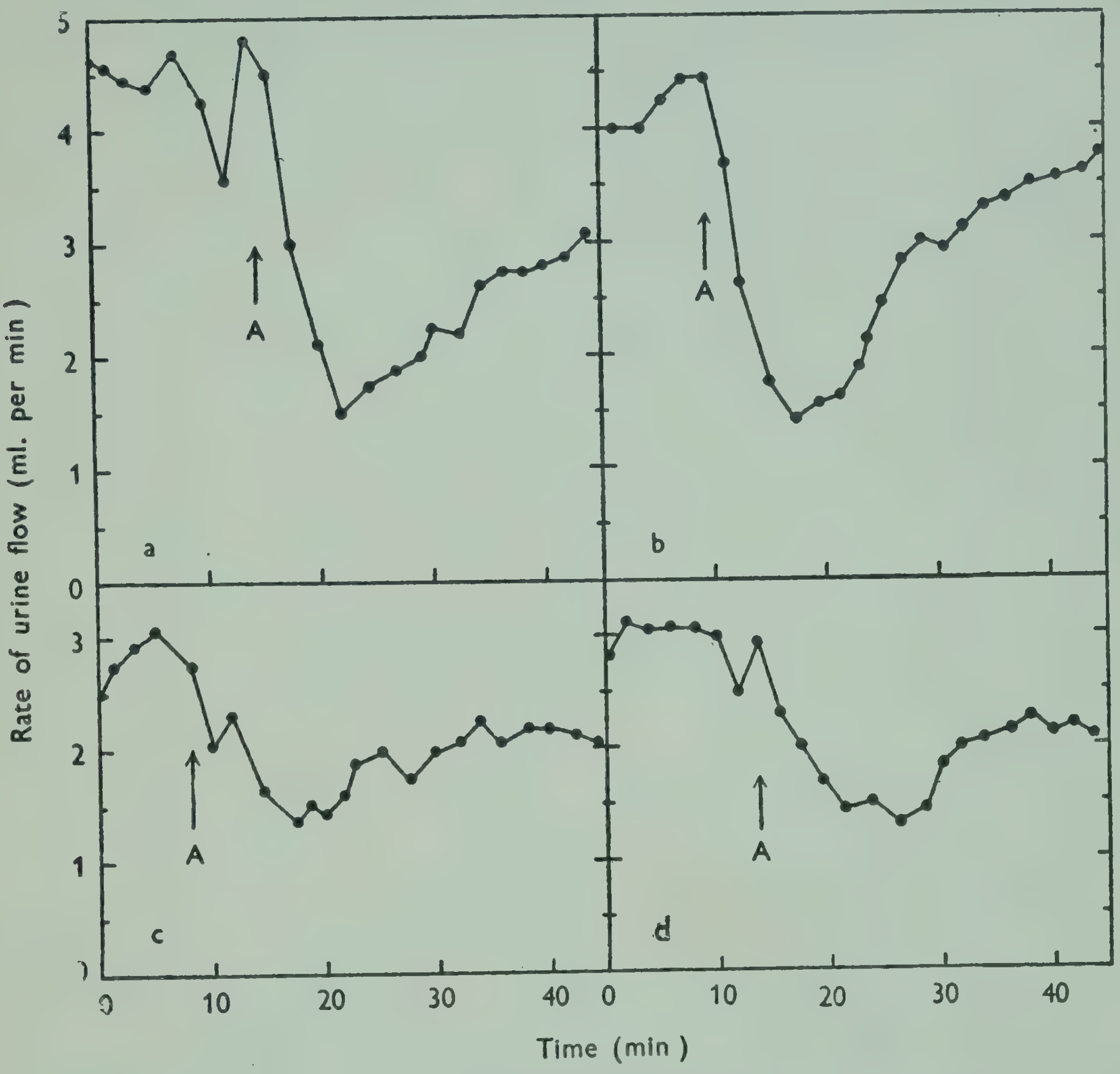


FIG. 34, 20

Urine flow (dog) after injections of sodium chloride solution and posterior pituitary extracts *a* and *b* 5 days before, *c* and *d* 11 days after, removal of pituitary. *a*: at A, injection of 21 ml. of 2·50 per cent. NaCl into right carotid in 20 sec; *b*: at A, 1·0 μU of postpituitary extract injected intravenously in 15 sec; *c*: at A, injection of 21 ml. of 2·50 per cent. NaCl into right carotid in 20 sec; *d*: at A, 0·1 μU postpituitary extract injected intravenously in 20 sec. (E. B. Verney (1947), *Proc. Roy. Soc.* B, **135,** 62).

The behaviour of the osmoreceptors depends on the nature of the substance responsible for the increase in the osmotic pressure of the blood. The injection of a hypertonic solution of urea rapidly (10 sec) into the carotid artery does not alter the flow of urine. Although a rapid injection of hypertonic glucose (Fig. 34, 19, A) is effective in reducing urine flow, a long-continued infusion of glucose sufficient to produce a rise of 80 mg./100 ml. in the carotid blood sugar has no effect on water elimination. The slow injection of sucrose produces the same secretion of antidiuretic hormone as a prolonged injection of sodium chloride of the same hypertonicity. To explain these findings Verney suggests that the osmoreceptors are freely permeable to urea, less freely permeable to glucose and relatively impermeable to sodium and sucrose. It has been calculated that an increase of osmotic pressure in the plasma of only 2 per cent. causes the pituitary to secrete about 1 μU (micro-unit) of ADH per second. Such an increase of osmotic pressure is equivalent to a rise in the concentration of Na$^+$ of 3 milli-equivalents per litre.

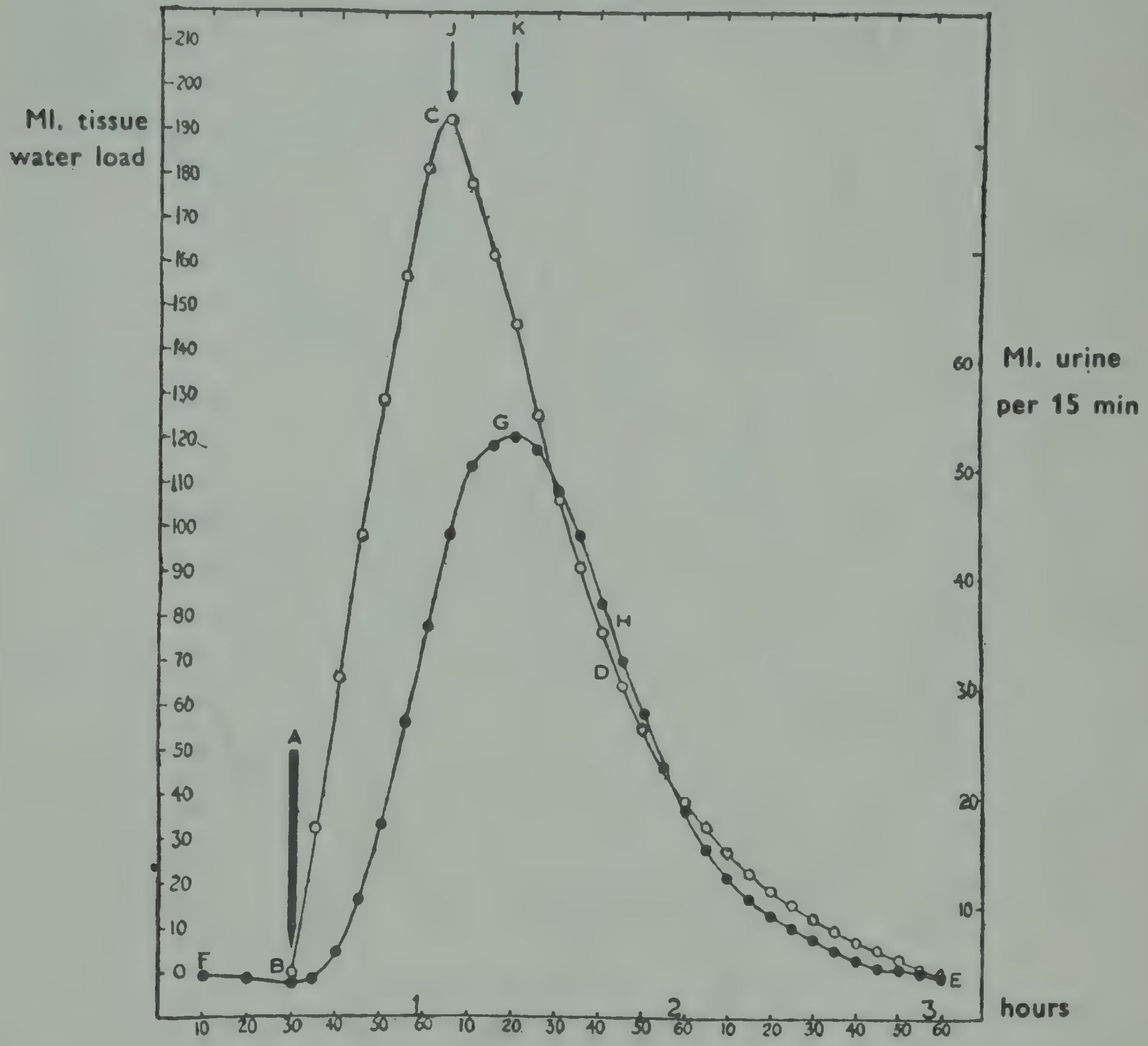


FIG. 34, 21

F, B, G, H, E = the mean of the mean rates of secretion obtained from nine dogs. The ordinate, in ml. per 15 minutes, is scaled on the right side. B, C, D, E represents changes in the water-load of the tissues with time. The ordinate of this curve is in ml. and is scaled on the left side. A = time of giving 250 ml. water. The interval JK between the peak of the water-load curve and the maximum rate of urinary secretion is 15 minutes (A. Klisiecki, M. Pickford, P. Rothschild & E. B. Verney (1933), *Proc. Roy. Soc.* B, **112**, 507).

If a long-continued intravenous infusion of pituitary extract is given to a normal dog at the rate of five milliunits per hour the urinary output is considerably reduced, but it returns to normal within half an hour of stopping the infusion. The antidiuretic hormone is apparently destroyed fairly quickly in the tissues or excreted in the urine; if this were not so water diuresis would be greatly delayed. Further, the recovery of the normal rate of urine flow when injection of hypertonic material ceases is probably due to the diminished output of hormone by the posterior lobe as the blood osmotic pressure declines, and partly to the excretion by the kidney of the antidiuretic substance.

Water Diuresis.—To study the effects of a reduction in the osmotic pressure of the blood and tissues, Verney gave dogs 250 ml. of water and by measuring the fluid remaining in the gut at various times after ingestion of the water he estimated the water load of the tissues, that is, the amount which had passed into the tissues. The peak of this load preceded the maximal urinary excretion by about fifteen minutes, whether the kidneys were intact or denervated (Fig. 34, 21). Presumably this effect occurs because the reduction in the concentration of sodium chloride and other osmotically active substances in the plasma causes a fall in the activity of the osmoreceptors and a fall in the output of the antidiuretic hormone; a little time must elapse to allow the hormone already present in the circulation to disappear before the reduction in the output of hormone from the pituitary can have its full effect. As Verney says, "water diuresis may be fitly and accurately described as a condition of physiological diabetes insipidus."

Suprarenal Cortex and Water Balance.—Suprarenal cortical hormones have at least two important effects on the elimination of water through the kidneys.

(*a*) The salt-retaining hormone aldosterone (p. 938) promotes the tubular reabsorption of water. Normally some 98 to 99 per cent. of the sodium in the glomerular filtrate is reabsorbed. A small diminution only in this fraction will lead to a fall in the sodium concentration of the extracellular fluid. This results in a hypotonic fluid and a loss of water therefore occurs. Finally the concentration of Na^+ in the extracellular fluid remains persistently low as in Addison's disease (Chap. 51).

(*b*) The diuretic response to water is impaired in patients with Addison's disease but is restored by the administration of cortisone (Chap. 51). Further, diabetes insipidus cannot be established in the experimental animal if the adenohypophysis is destroyed. Adrenal cortical hormones are probably required to provide an essential background for the regulation of the excretion of water, though they do not control it directly.

EFFECT OF EMOTION ON URINARY OUTPUT.—Verney found that if his dogs were exercised they showed an inhibition of urine production to a variable extent beginning about two minutes after the exercise commenced. Since this effect disappeared if the animals were repeatedly exercised it was not the result of the exercise itself but was more likely to be due to the emotional stress associated with it, since a loud noise or electrical stimulation of the skin also inhibited urine flow. This phenomenon was observed even when the kidneys and suprarenals were denervated and was presumed to be due to the release of antidiuretic hormone from the posterior pituitary in response to impulses from the hypothalamic nuclei (Fig. 34, 17). If the posterior pituitary was removed the response to emotion was reduced to about one-twentieth (compare Fig. 34, 20).

Emotional stress in man also reduces the output of urine (p. 610). This could be attributed to an outpouring of adrenaline from the suprarenal medulla or to the discharge of antidiuretic hormone. Adrenaline produces a marked but transient reduction in urinary output by constricting the efferent glomerular arterioles (p. 610). These effects on diuresis must not be confused with emotional alterations of the micturition reflexes; emotion often causes frequency of micturition without actual increase in the volume of urine passed.

The Kidney and Œdema

There is no doubt that the retention of salt by the kidney is an important factor in the production of œdema and a salt-free diet is one of the most valuable therapeutic measures for its relief. The retention of sodium is followed by retention of water, so that the osmotic pressure remains normal at the expense of an accumulation of water in the tissues and an increased blood volume (*hypervolæmia*). The retention of water finally becomes evident as œdema. Because similar changes can be produced experimentally by large doses of sodium bicarbonate or sodium bromide, but not by potassium or ammonium chloride, it is concluded that sodium is the important factor.

An excessive retention of sodium chloride and water occurs in heart failure and has been considered to be a primary factor in the production of the œdema. Heart failure results in increased venous pressure and therefore leads to raised filtration pressure in the capillaries, accumulation of fluid in the tissues and a reduction in circulating blood volume. Salt and water are retained by the kidney and the normal blood volume is restored and even exceeded. This conception of cardiac œdema emphasizes the importance of a raised venous pressure. A second theory draws attention to the fall in renal blood flow in many cases of cardiac failure and the consequent retention of salt and water. This leads to hypervolæmia, increase in venous pressure, and therefore capillary filtration pressure, and finally to œdema. According to this view the raised venous pressure is a secondary phenomenon. At present it is impossible to decide between the two conceptions of the mechanism of œdema formation in cardiac failure.

REGULATION OF THE HYDROGEN-ION CONCENTRATION OF THE BLOOD

The kidney, as we have seen, controls, under the direction of the hypothalamus, the osmotic pressure of the tissue fluids. Another important aspect of its homoiostatic activities is the regulation of the hydrogen-ion concentration of the body fluids. In acidosis (p. 643) the kidney usually produces an acid urine while in alkalosis (p. 643) the urine becomes alkaline. In both cases the alteration in renal function is compensatory and may result in the maintenance of the normal blood pH.

Although the part played by the kidney in the regulation of acid-base balance is only just beginning to be understood, it is already clear that it involves more than variations in the excretion of hydrogen ion. For example, the formation of ammonia from amino-acids and glutamine in the distal tubules is a mechanism of considerable importance especially when large amounts of fixed acid have to be eliminated. For the successful maintenance of acid-base equilibrium by the kidney an adequate urine flow is essential and the importance of giving large amounts of fluid to a patient with acidosis or alkalosis cannot be over-emphasized.

The alteration in pH as the fluid passes down the nephron has been studied by Ellinger by injecting either a frog or a rat with fluorescein which, on illumination with ultraviolet light, fluoresces deep green at pH 9 and canary-yellow at pH 3. Since the fluid in the glomerular capsules is greenish, like blood plasma, whereas lower down the tubule it becomes progressively more yellow, it may be concluded that a gradual reduction in pH has occurred in transit through the tubule. Acidification normally occurs mainly in the distal convoluted tubule, although, when large doses of ammonium chloride are given, preliminary reduction in pH occurs in the proximal tubule.

The main buffering substances of the blood which appear in the glomerular filtrate are phosphates and bicarbonates. The urine, however, normally contains very little bicarbonate, since this is almost completely reabsorbed by the cells of the tubules from the glomerular filtrate. Plasma and urine both contain the ions $H_2PO_4^-$ and $HPO_4^=$. In plasma the ratio $H_2PO_4^-$: $HPO_4^=$ is about 1: 5, but in urine it is usually 9: 1 and may rise as high as 50: 1 when large amounts of acid have to be eliminated. In explanation of these observations it is suggested that an ionic exchange occurs in which the following reaction takes place in the proximal tubules:

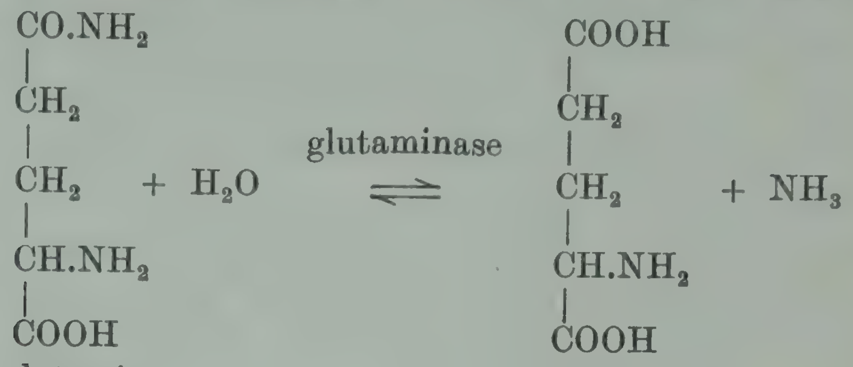
$$Na_2HPO_4 + H_2CO_3 = NaH_2PO_4 + NaHCO_3$$

The dihydrogen phosphate is then excreted while the bicarbonate is absorbed. After absorption it may provide base for the reconstitution of other buffer salts by yielding carbon dioxide which is removed by the lungs.

Pitts believes that not more than one-third of the acidity of urine can be accounted for by reabsorption of bicarbonate and phosphate and suggests that the tubules secrete hydrogen ions into the tubular fluid and that carbonic anhydrase, present in the cortex of the kidney, is involved in this reaction in much the same way as it is concerned in the production of hydrogen ions in the stomach (p. 240). Since sulphonamides inhibit carbonic anhydrase, patients under treatment with these drugs may have an impaired ability to secrete hydrogen ions into the kidney tubules and their urine may be less acid than normal.

It is sometimes desirable clinically to alter the pH of the urine and a number of drugs have been used for this purpose. The urine can be made alkaline by the ingestion of sodium bicarbonate, but large doses must be given at short intervals since the excess of bicarbonate in the plasma is rapidly excreted in the urine. At the same time the dietary chloride should be restricted, as a high chloride intake diminishes the bicarbonate excretion. The urine can also be rendered alkaline by the administration of very large doses of sodium citrate. This substance is metabolized to give carbon dioxide and sodium bicarbonate which is excreted in the urine. The urine can be made more acid by the administration, in doses of three to six grams a day, of ammonium chloride which, in the body, in effect yields ammonia and hydrochloric acid. The ammonia combines with carbon dioxide or bicarbonate to yield urea which has no effect on urinary pH, while the hydrochloric acid, which tends to increase the hydrogen-ion concentration, is eliminated in the urine.

When large quantities of acid substances are being produced in metabolism they can be neutralized by ammonium salts synthesized by the kidney which contains the enzyme *glutaminase.* This breaks down glutamine with the formation of glutamic acid and ammonia.



$$\begin{array}{l} CO.NH_2 \\ | \\ CH_2 \\ | \\ CH_2 \\ | \\ CH.NH_2 \\ | \\ COOH \end{array} + H_2O \underset{}{\overset{\text{glutaminase}}{\rightleftharpoons}} \begin{array}{l} COOH \\ | \\ CH_2 \\ | \\ CH_2 \\ | \\ CH.NH_2 \\ | \\ COOH \end{array} + NH_3$$

glutamine glutamic acid

The glutamine provides a readily available supply of ammonia, which is released by the glutaminase present in the kidney cells. When large amounts of acid have to be eliminated ammonia is excreted into the lumen of the tubules and base, largely sodium, is reabsorbed to maintain the electrolyte balance. By this mechanism as much as 50 to 60 ml. of N acid can be neutralized per day. Ammonia does not appear to be produced from urea to any important extent.

NEPHRECTOMY

The removal of one kidney is quite compatible with a normal active life provided that the remaining kidney, which increases somewhat in size after the operation, remains healthy. Some two-thirds to three-quarters of the total renal substance must be removed or destroyed before renal failure occurs. It has occasionally happened that one kidney has been removed when the remaining kidney, unknown to the operator, has been functionless. In these circumstances the substances normally excreted accumulate in the body and the patient dies in about a week. The concentration of the urea in the blood increases from the normal of 20 to 30 mg. per 100 ml. up to as much as 900 mg. per 100 ml., and there is retention of other nitrogenous bodies such as uric acid and creatinine with severe acidosis. The patient looks ill and suffers from vomiting and diarrhœa, and later breathlessness. At first he is listless, then he becomes confused and progressively more drowsy, until he lapses into coma and finally dies. The blood pressure is never raised. The mechanism of this condition, called *uræmia,* is not clear, but it cannot be accounted for solely by the retention of urea and uric acid since neither is toxic even in large doses.

RENAL FUNCTION IN THE NEW-BORN

The glomeruli are all formed in the new-born infant but they are much smaller and more crowded together than in the adult, and the remainder of the nephron is undeveloped. The glomerular capsule which covers the tuft of capillaries consists of tall columnar cells in fœtal life, of cubical cells in the new-born and of flattened cells in the adult; it is probable that the greater thickness of the glomerular membrane accounts for the relatively low glomerular filtration rates found in early life. In new-born rats the nephrons in the peripheral parts of the cortex do not excrete trypan blue, an ability that is acquired only when the brush border begins to develop.

That regulation of the internal environment is carried on *in utero* chiefly by the placenta is indicated by the fact that apparently normal children may be born with functionless kidneys. The steady rise of urea and uric acid in the liquor amnii during pregnancy suggests that the kidneys normally begin to function before birth, although only to a small extent. The urine of babies examined immediately after birth is generally acid with a small amount of urea and a low osmotic pressure; albuminuria is not uncommon in the first five days of extrauterine life.

Immediately after birth the body becomes dependent on the kidney for elimination of nitrogenous end-products of metabolism and to a large extent also for the regulation of its internal environment. The plasma protein level is low in infancy and, therefore, the colloid osmotic pressure of the plasma is lower than in the adult, a fact which favours the filtration of fluid through the glomerular capillaries as well as through capillaries in all parts of the body. The resulting tendency to accumulation of extracellular fluid and development of œdema is, however, offset by the low fluid intake which ordinarily occurs during the first few days of extrauterine life, and the volume of the extracellular fluid actually tends to fall; at the same time the blood urea concentration rises. In a few days, as the fluid intake increases, the extracellular fluid volume rises and the blood urea falls. Another factor accounting for the relatively large extracellular fluid volume of the human infant, especially if it is premature, is the low renal clearance of sodium and chloride.

In babies a few days old glomerular filtration rates measured by inulin clearance are about 30 ml./sq.m/min compared with about 75 ml./sq.m/min in the adult. Tubular function is also much less efficient than in adults for the renal clearances of diodone and *p*-aminohippurate are only about one-tenth of adult values when allowance is made for the differences in surface area. Filtration fractions (p. 607) of these substances are high, a finding that suggests that tubular function is less developed than glomerular function in the human infant. Urea clearances are also low in infancy. All these values increase to adult level towards the end of the first year.

In new-born children the volume of urine passed per day is of the order of 20 ml. and this increases only when the intake of fluid becomes greater as the maternal milk supply improves. Water given to babies up to three months of age is excreted very slowly and this is true of other new-born mammals. The specific gravity and osmotic pressure of the urine are low. Not only has the administration to very young babies of posterior pituitary extracts containing antidiuretic hormone very little effect in making the urine more concentrated, but the posterior lobe of the pituitary gland in very young animals contains very little antidiuretic hormone. The production of antidiuretic hormone by the pituitary gradually increases as the kidney develops.

Infants easily become dehydrated and the administration of fluids readily produces œdema. The low phosphate clearances and low phosphate excretion make compensation for acidosis inadequate. The limited powers of renal function explain how the chemical composition of the plasma in the new-born may be quickly upset by disease and why the condition so easily becomes irreversible. Some of the pathological changes in children which occur as the result of diarrhœa or vomiting are often, and probably correctly, ascribed to renal inadequacy. The kidney may not, however, be actually abnormal in these conditions for, as we have seen, the normal kidney of the human infant, with its limited powers of elimination of water and solutes, is inadequate by adult standards.

MICTURITION

The urine is conveyed from the kidneys in the ureters which by passing obliquely through the thick muscular wall of the bladder acquire a valvular opening so that urine is prevented from regurgitating into the ureter when the intravesical pressure is raised during micturition. In this way the kidney and ureters are protected from the variations in pressure which occur physiologically in the bladder. The urine is carried along the ureters by peristaltic waves which travel down to the bladder about three times a minute, each wave sending a little spurt of urine into the bladder. This can readily be observed in man through an instrument called a cystoscope inserted along the urethra into the bladder after the urine has been coloured by a dye injected intravenously and excreted by the kidney.

The bladder is lined by transitional epithelium firmly adherent to the wall of smooth muscle fibres, the *detrusor* muscle, which is supplied by the hypogastric nerves via the hypogastric ganglia and by the pelvic nerves through their ganglia on or near the bladder wall. At the base of the bladder the smooth muscle fibres pass around the urethral opening loopwise to form the *internal sphincter* (sphincter vesicæ) which is supplied by hypogastric nerves (lumbar or sympathetic outflow) and the pelvic nerves (sacral or parasympathetic outflow) (Fig. 34, 22). The *external sphincter* (sphincter urethræ) surrounding the membranous urethra is composed of two striated muscles, the compressor urethræ and the bulbocavernosus, and is under voluntary control through the pudendal nerves which are somatic nerves. This arrangement should be compared with that for the rectum and colon (p. 277). Sensory fibres subserving pain (muscle pain) run in the sympathetic and may, judging by the effects of cord injury or partial anæsthetization of the cord in man, run up as high as the 9th thoracic segment before entering the cord. Mucous membrane pain is conveyed in the parasympathetic. Since both sympathetic and parasympathetic fibres pass through the hypogastric ganglion (Fig. 34, 22), section of all nerves passing from the bladder to the ganglion produces complete anæsthesia of the bladder.

In the empty bladder at rest the sphincters are tonically contracted so that the urine from the ureters gradually expands the bladder without, however, causing the intravesical pressure to rise very much (Fig. 34, 23). This is usually explained by supposing that the detrusor muscle relaxes as the bladder expands but, while this may be true, the small rise in pressure in the bladder is partly the result of the elastic properties of the bladder wall, since it can be shown that in any elastic bag the pressure, except in the early stages of filling, is almost independent of the volume of the contents.

The tension in the walls is, however, proportional to the contained volume. If the pressure is increased suddenly by introducing, say, 200 ml. of fluid through a catheter into the bladder, rhythmic contractions are set up, though they quickly die out. When the bladder contains over 400 ml. the detrusor muscle contracts rhythmically, and if it contains over 600 ml. the contractions become powerful enough to raise the pressure to 100 cm of water, and may be painful. If voluntary efforts at restraint are made (Fig. 34, 24) the pressure may fall for a short time, but when the pressure exceeds 100 cm of water again discomfort becomes acute and micturition inevitably occurs. The level of pressure at which the desire to micturate occurs depends partly on the rate of filling—if this is slow the bladder can accommodate more fluid—and partly on the irritability of the nerve endings in the bladder and their central connexions. Afferent impulses travel in the pelvic nerves (and probably in the hypogastric nerves) and emptying occurs by (1) impulses from the pelvic nerves which cause the detrusor muscle to contract and the internal sphincter to relax, and (2) a reduction in the number of impulses in the hypogastric nerves so that the internal sphincter relaxes. The pelvic nerves are certainly the most important in emptying, because section of both pelvic

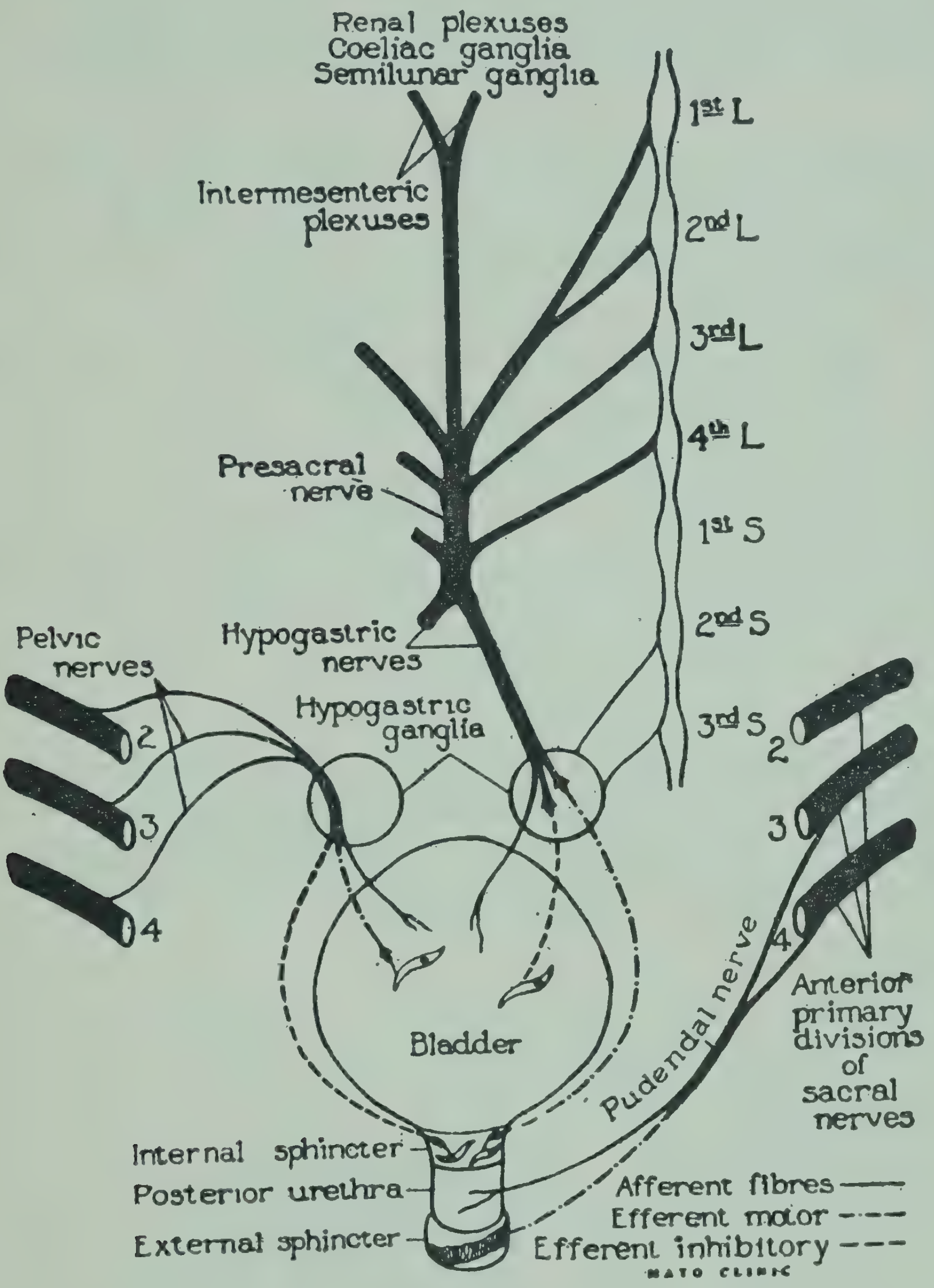

Renal plexuses
Coeliac ganglia
Semilunar ganglia
1st L
Intermesenteric plexuses
2nd L
3rd L
4th L
Presacral nerve
1st S
2nd S
Pelvic nerves
Hypogastric nerves
Hypogastric ganglia
3rd S
2
3
4
2
3
4
Anterior primary divisions of sacral nerves
Pudendal nerve
Bladder
Internal sphincter
Posterior urethra
External sphincter
Afferent fibres
Efferent motor
Efferent inhibitory
MAYO CLINIC

FIG. 34, 22

Diagram of innervation of the bladder and sphincters in man (J R. Learmonth (1932), *Amer. J. Surgery*, **16**, 270).

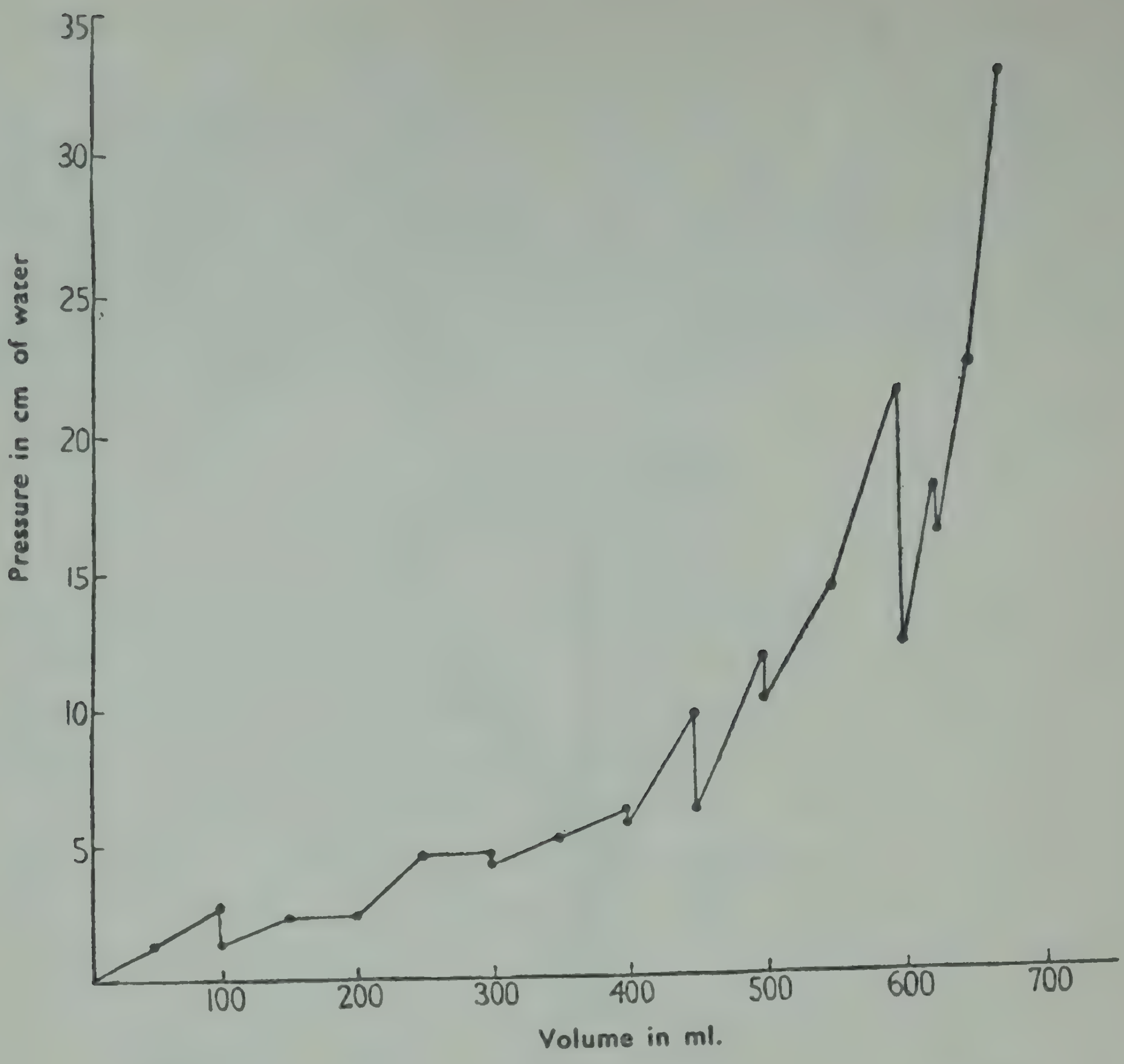


FIG. 34, 23

Graph of the relationship of the volume of the vesical contents to their pressure in the course of distension of the human bladder. Each vertical fall of pressure indicates the degree of adaptation at a constant volume (D. Denny-Brown & E. Graeme Robertson (1933), *Brain*, **56**, 158).

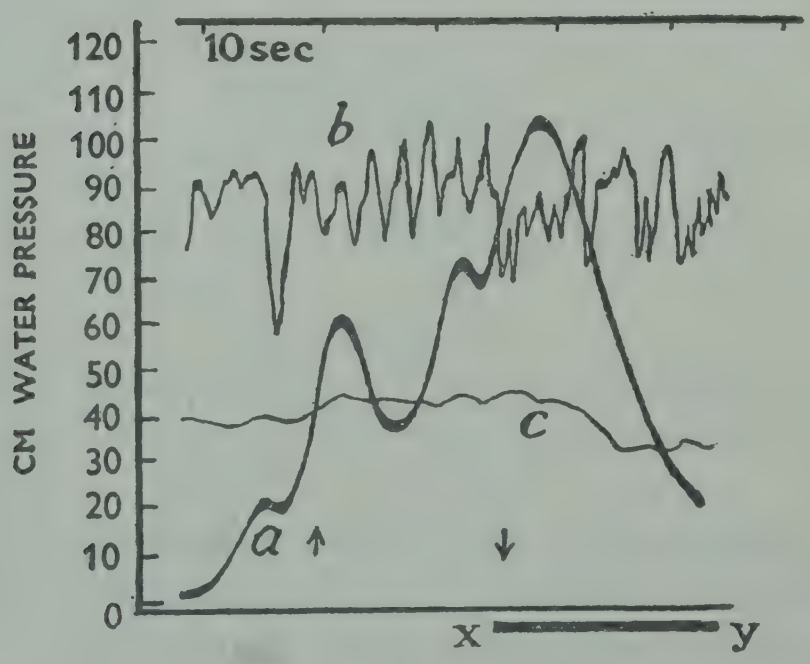


FIG. 34, 24

Records of bladder pressure (heavy curve, *a*), intrarectal pressure (*c*) and movements of the abdominal wall (*b*) a few seconds after the beginning of an attempt to micturate. At the first arrow the sphincters opened. A powerful effort to restrain micturition was made from *x* to *y* and the sphincters closed at the second arrow. There is a considerable fall of vesical pressure during the period *x*–*y* (D. Denny-Brown & E. G. Robertson (1933), *Brain*, **56**, 163).

nerves in the cat causes immediate paralysis of micturition although partial recovery may eventually take place; the detrusor muscle shows only feeble contractions so that true micturition can scarcely be said to occur. This statement is probably true also for man. When, in man or cat, the hypogastric nerves alone are divided the interference with evacuation of urine is not very serious. Impulses in the hypogastric nerves maintain the tone of the internal sphincter and by reduction of this tone the internal sphincter is allowed to open; these nerves exercise only a small and variable influence on the detrusor, although it is likely that they are responsible for the progressive relaxation of this muscle as the organ gradually fills.

The effect of spinal anæsthesia in man is to relax slightly the internal sphincter and the ureteric orifices. In transverse lesions of the spinal cord in man which result in paraplegia (p. 829) the internal sphincter is at first contracted and the detrusor muscle is flaccid; urine accumulates and the bladder distends, but eventually the pressure rises high enough to force urine through the sphincter. After small amounts of urine have been voided the pressure is somewhat reduced and urine accumulates again in the bladder until the process repeats itself. This is *retention incontinence.* After a month or so it will be found that micturition occurs automatically, usually at hourly intervals, by contraction of the detrusor muscle. The capacity of the bladder is then smaller than normal and a small increase in its contents causes a considerable rise of pressure. Provided the sacral cord and its nerve connexions with the bladder are intact the bladder reflexes in man with a transverse lesion of the spinal cord occur in the usual way, except that they are not under restraint from the higher centres which normally exercise a constant inhibition on the detrusor muscle. Destruction of the cauda equina in man leaves the external sphincter flaccid and interferes with the reciprocal activity of the detrusor and internal sphincter. The latter gives way at much lower pressure than normal when spontaneous contractions of the bladder occur. There is a rudimentary micturition mechanism since, although the contractions of the bladder are not very vigorous, they can be reinforced by voluntary contractions of the abdominal muscles.

In all these descriptions it will have been noted that the internal sphincter has a reciprocal relationship with the detrusor, the former relaxing when the latter contracts. The external sphincter, although made of striated muscle, cannot be relaxed voluntarily although it may be consciously contracted. Efforts to micturate do not produce a relaxation of the external sphincter unless the intravesical pressure is above a threshold value of from 20 to 40 cm of water. Inhibition of the sphincters occurs only when the detrusor muscle is contracting, the internal sphincter relaxing a little before the external. The stream of urine can, however, be stopped by a voluntary contraction of the external sphincter.

Other reflexes contributing to the emptying of the bladder may arise when urine passes along the urethra and distension of the bladder causes reflexly a relaxation of the urethra.

Although the micturition reflexes can be carried out in man by the cord alone, the bladder and its sphincters are normally influenced by the higher centres. A voluntary effort to micturate evokes powerful contractions of the bladder after a very short latent period and voluntary restraint of micturition results in an inhibition of these contraction waves (Fig. 34, 24). Normal micturition depends entirely on active contraction of the bladder; voluntary control, that is the ability to delay or start micturition when desired, should be thought of as inhibition or removal of inhibition respectively of the normal reflex response to distension which is dependent on the lower parts of the spinal cord. According to this view the infant develops voluntary control of micturition by acquiring the power to inhibit the spinal reflexes.

Certain associated movements occur in micturition which are not regarded as essential parts of the process. An attempt to micturate may be accompanied by relaxation of the perineal muscles (but not of the sphincters), contraction of the abdominal muscles and occasionally movements of the lower colon and rectum. Voluntary restraint is associated with a contraction of the perineal muscles and closure of the external sphincter.

REFERENCES

Addis, T. & Oliver, J. (1931). *The Renal Lesion in Bright's Disease.* New York: Hoeber.

Berglund, H. & Medes, G. (1935). *The Kidney in Health and Disease.* Philadelphia: Lea & Febiger.

Bradley, S. E. (Ed.) (1950). *Renal Function.* New York: Josiah Macy Jr. Foundation.

Cushny, A. R. (1926). *The Secretion of Urine.* 2nd ed. London: Longmans.

Denny-Brown, D. & Graeme Robertson, E. (1933). The Physiology of Micturition. *Brain,* **56,** 149–190.

Gruber, C. M. (1933). The autonomic innervation of the genito-urinary system. *Physiol. Rev.* **13,** 497–609.

Langworthy, O. R., Kolb, L. C. & Lewis, L. G. (1940). *Physiology of micturition.* Baltimore: Williams and Wilkins.

Learmonth, J. R. (1931). A contribution to the neurophysiology of the urinary bladder in man. *Brain,* **54,** 147–176.

McCance, R. A. (1950). Renal physiology in infancy. *Amer. J. Med.* **9,** 229–241.

O'Connor, W. J. (1947). Control of urine secretion in mammals by pars nervosa of pituitary. *Biol. Rev.* **22,** 30–53.

Pickering, G. W. (1955). *High Blood Pressure.* London: Churchill.

Pickford, M. (1952). Antidiuretic substances. *Pharmacol. Rev.* **4,** 254–283.

Pitts, R. F., Sullivan, W. J. & Dorman, P. J. (1954) in Ciba Foundation Colloquium on *The Kidney.* London: J. & A. Churchill.

Prather, G. C. (1949). *Urological Aspects of Spinal Cord Injuries.* Oxford: Blackwell.

Richards, A. N. (1934–35). Urine formation in the amphibian kidney. *Harvey Lect.* **30,** 93–116.

Robinson, J. R. (1954). *Reflections on Renal Function.* Oxford: Blackwell.

Shannon, J. A. (1939). Renal tubular excretion. *Physiol. Rev.* **19,** 63–93.

Smith, H. W. (1937). *The Physiology of the Kidney.* Oxford University Press.

Smith, H. W. (1939–40). Physiology of the renal circulation. *Harvey Lect.* **35,** 166–189.

Smith, H. W. (1943). *Lectures on the Kidney.* Lawrence, Kansas: University of Kansas.

Smith, H. W. (1951). *The Kidney: Structure and Function in Health and Disease.* Oxford University Press.

Symposium: Histochemistry and the elucidation of kidney structure and function (1955). *J. Histochem. Cytochem.* **3,** 243–312.

Verney, E. B. (1929). Polyuria. *Lancet,* **1,** 539–546, 645–651, 751–756.

Winton, F. R. (1951–52). *Harvey Lect.* **47,** 21–44.

CHAPTER 35

THE FLUIDS OF THE BODY

IT is only in the liver and spleen that there are actual gaps in the endothelial wall of the capillaries, allowing the blood to come into direct contact with the tissue cells. In all other organs the cells are bathed by *interstitial fluid*, which occupies the space between the blood and the tissue cells containing *intracellular fluid*. This space (the *interstitial space*, Fig. 35, 1) is bounded by the endothelial walls of the capillaries and the cell membranes of the tissue cells. Interstitial fluid contains about 0·1 per cent. of protein whereas the plasma has 7 per cent. In other respects their chemical constitution is very similar and there is a continual rapid exchange of material between them. It is therefore justifiable to think of the protein-free part of the plasma together with lymph, and the fluid in the interstitial space as the *extracellular fluid* (Fig. 35, 1). Its chemical constitution is relatively easy to study by examining deproteinized blood withdrawn from a vein. Because samples are easily obtained a good deal is known about the amount and composition of extracellular fluid in health and disease. Much less is known about intracellular fluid which can be studied only indirectly.

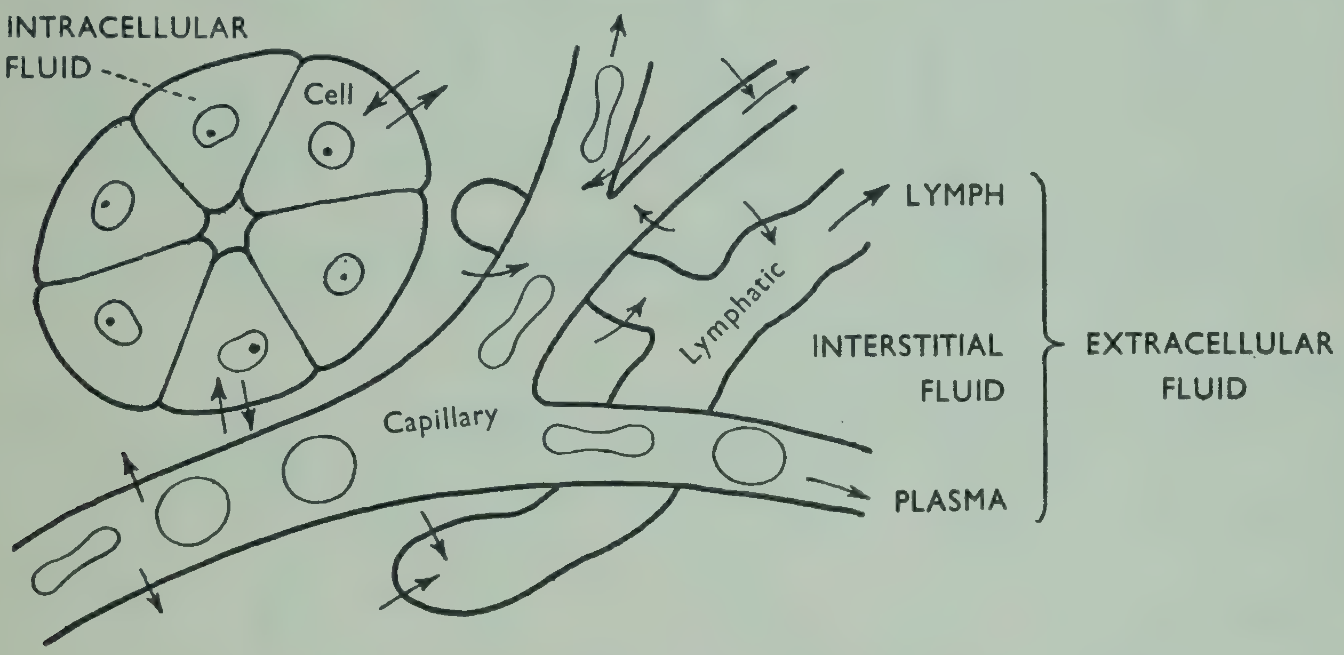
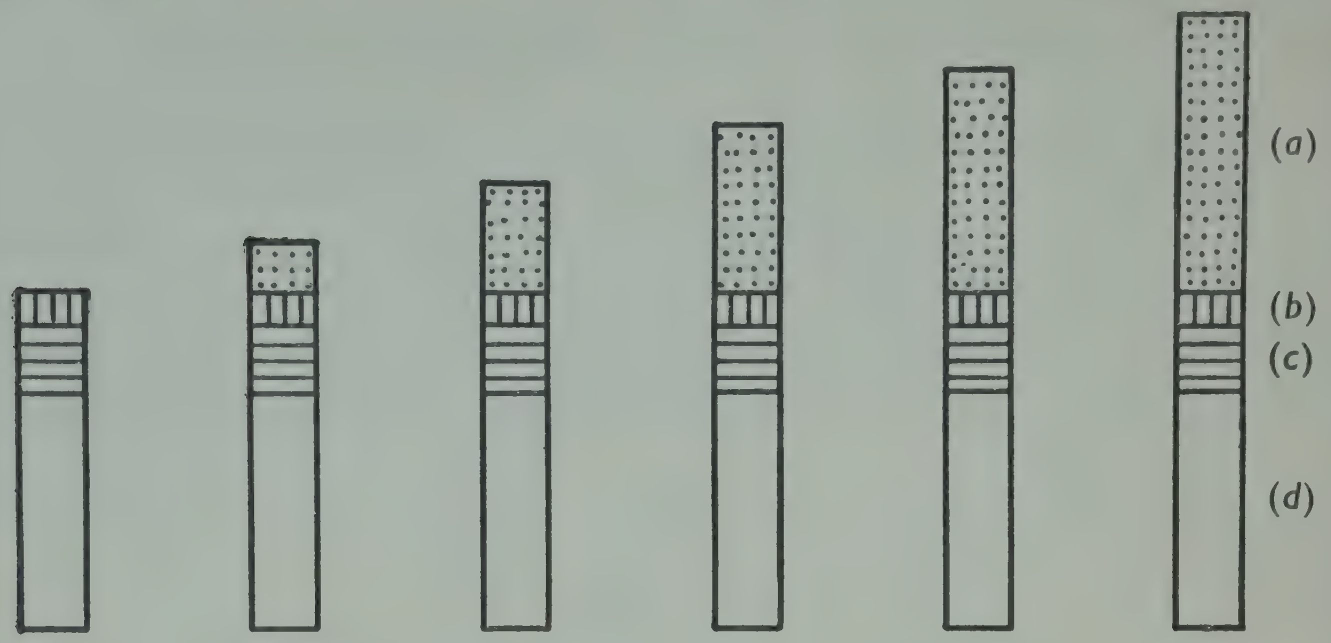


Fig. 35, 1

Diagram to show the main subdivisions of the body fluid. The term extracellular fluid includes interstitial fluid, lymph and the protein-free fraction of plasma. (*After* Cowdry, *Textbook of Histology*, 1938, p. 36. London: Kimpton.)

TOTAL BODY WATER

The amount of water in the human body is remarkably constant (approximately 42 litres). When total body water is considered as a percentage of body weight, however, then as the body weight increases because of the deposition of more and more fat containing little water, the percentage of the body weight occupied by water falls from some 60 per cent. to 35 to 40 per cent. (Fig. 35, 2). By determining the specific gravity of the body in a large number of subjects and thus calculating their fat content Behnke was led to the conception of a basic body structure or *lean tissue mass*, more or less constant in composition, containing 70 per cent. of water, 20 per cent. of solids and an irreducible minimum of 10 per cent. of fat. Similar values for lean tissue mass have also been found in such dissimilar species as mice, guinea-pigs and monkeys, although the size and fat contents of the animals differed widely.



	BASIC STRUCTURE	NORMAL ADULT MALE	INCREASING DEGREES OF FATNESS			
TOTAL BODY WEIGHT						
lb.	132	154	176	190	220	242
kg.	60	70	80	90	100	110
STORAGE FAT						
kg.	0	10	20	30	40	50
% Body Weight	0	14	25	33	40	45
TOTAL BODY WATER						
Litres	42	42	42	42	42	42
% Body Weight	70	60	54	45	42	38

Key.—(*a*) Excess or "Storage" fat; (*b*) Essential or structural fat; (*c*) Lean tissue solids; (*d*) Water.

FIG. 35, 2

The relation between body fat and body water (A. W. Wilkinson (1955), *Body Fluids in Surgery*. Edinburgh: E. & S. Livingstone).

The total body water can be measured in man by injecting into a vein a known quantity of a substance which fairly rapidly becomes evenly distributed throughout the body water. The concentration of the injected substance in the plasma after equilibration is then a measure of the extent of its dilution from which the total amount of water through which it has diffused can easily be calculated. Deuterium oxide (heavy water) (p. 286), tritium oxide (p. 287), and the drug antipyrine are suitable substances. The water content of the body is relatively high in newborn infants (77 per cent. of body weight) and falls gradually after the age of fifty. In the normal adult male the water content of the body is about 60 per cent., but in the female it is only 55 per cent. because of the greater proportion of fat.

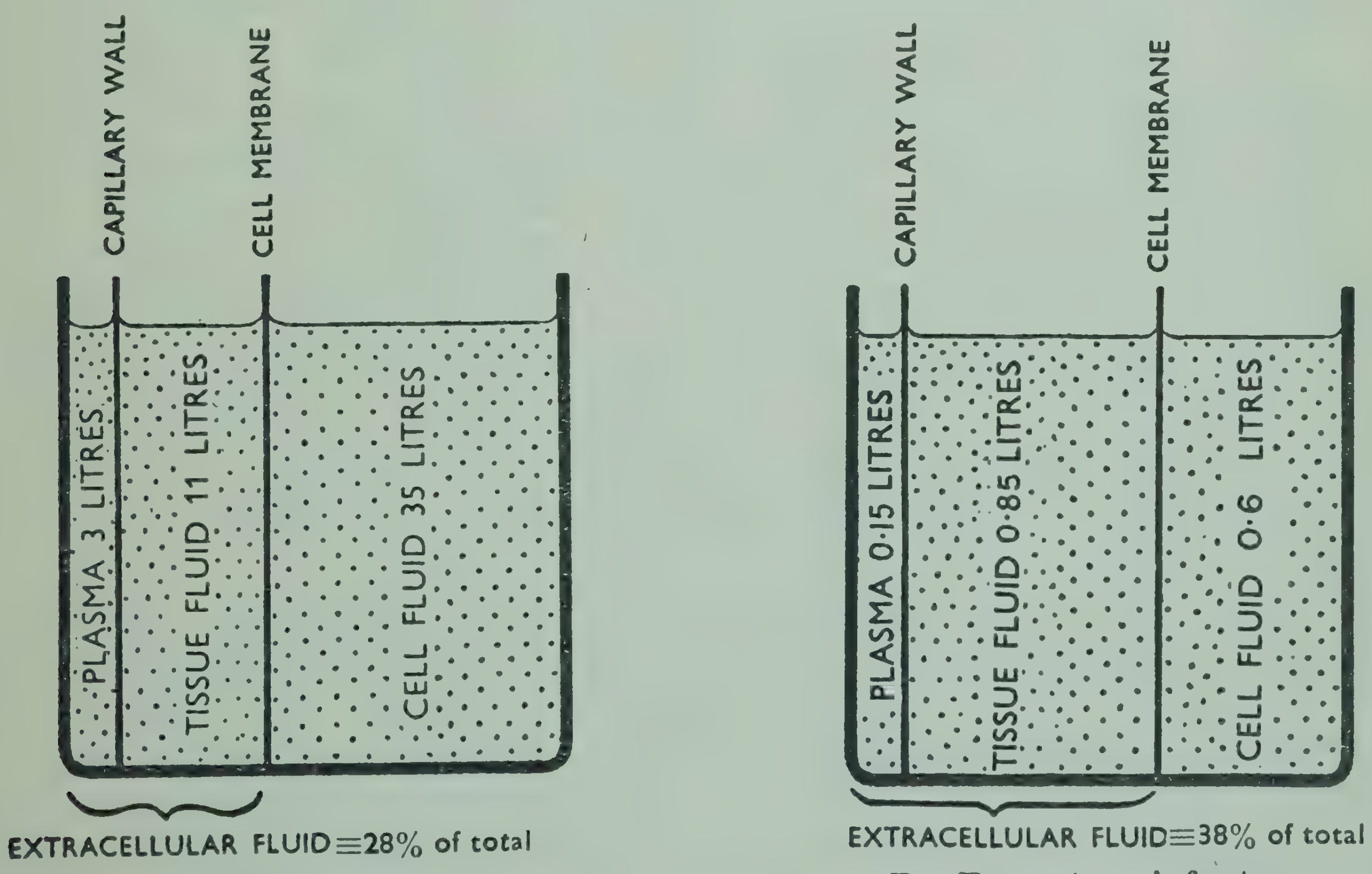
Apart from a small amount (10 per cent.) present in bone the total amount of water in the body is distributed approximately as shown in Table 35, 3. The distribution is indicated diagrammatically in Fig. 35, 4,

TABLE 35, 3

Distribution of Water in the Human Body

	Percentage of Total Body Water	Percentage of Body Weight	Volume (litres) in 70 kg. adult
Total body water . .	..	60·0	42
Intracellular fluid . .	70	42·0	30
Extracellular fluid. .	28	16·8	12
Interstitial fluid. .	21	12·6	9
Plasma . . .	7	4·2	3

which shows also the big difference between the partition of the total body water in the adult and the premature infant.


CAPILLARY WALL
CELL MEMBRANE
PLASMA 3 LITRES
TISSUE FLUID 11 LITRES
CELL FLUID 35 LITRES
EXTRACELLULAR FLUID≡28% of total
CAPILLARY WALL
CELL MEMBRANE
PLASMA 0·15 LITRES
TISSUE FLUID 0·85 LITRES
CELL FLUID 0·6 LITRES
EXTRACELLULAR FLUID≡38% of total

A. Adult man

B. Premature infant

FIG. 35, 4

Diagrammatic representation of body fluids in adult man (*After* H. L. Marriott (1947), *Brit. med. J.* i, 245) and the premature infant (*After* R. A. McCance (1950), *Amer. J. Med.* **9**, 229).

The extracellular fluid volume can be measured by injecting into a vein a known amount of a substance that can distribute itself rapidly through the extracellular space—that is the interstitial space plus the plasma volume. After time has been allowed for equilibration the concentration of the injected material in the plasma permits calculation of the volume of fluid in which it has been diluted. Sodium thiocyanate and inulin (p. 17) have both been used for this purpose.

The volume of intracellular fluid is obtained by the difference between total body water and extracellular fluid.

The determination of plasma volume has already been described (p. 376).

Composition of Body Fluids

To speak strictly, the term body fluids should comprehend all the fluids in the body. As ordinarily used, however, along with intra- and extracellular fluid, it includes the pleural, peritoneal and pericardial fluids whose composition closely resembles that of interstitial fluid and lymph, while excluding cerebro-spinal fluid, synovial fluid, and the fluids of the eye, inner ear, and labyrinth, all of which differ considerably from one another as well as from intra- and extracellular fluids. The lymph is considered separately later (p. 638).

Although we have spoken of the body water as being distributed between three spaces, the intracellular space, the extracellular space and the vascular compartment, there is a rapid and continuous exchange of fluid between them. The cell membranes are freely permeable to water but selectively to electrolyte ions. The movement of water between the cells and the interstitial space is therefore controlled by a dynamic equilibrium, water passing freely from the fluid of lower osmotic pressure to that of the higher. The membrane separating the vascular from the interstitial compartment is, however, freely permeable to both water and electrolyte ions and the equilibrium between plasma and interstitial fluid is maintained by the osmotic pressure of the plasma protein.

From what has been said in Chapter 4, it is obvious that the body water must contain substances dissolved in it, and that under basal conditions at any rate the intracellular and interstitial fluids must be practically isotonic (isosmotic) with blood plasma, that is to say must contain concentrations of electrolytes equivalent to 310 milliosmoles (p. 29) per litre. We must therefore consider the ionic constitution of the fluid in the three main body spaces.

In Fig. 35, 5 the composition of each of the three chief subdivisions of the body fluid is described in terms of a double column representing the cations on the left and the anions on the right. The height of each subdivision is proportional to the milliequivalence per litre of water. The figure thus shows that the plasma is a somewhat more concentrated solution than interstitial fluid and that the intracellular fluid is more concentrated still. These differences in ionic concentration are determined by Donnan equilibria (p. 36) due to the fact that both plasma and cell fluid contain protein which cannot pass through the membranes (capillary wall or cell membrane) separating them from the interstitial fluid.

It is clear from Fig. 35, 5 that the sodium ion is the principal cation of extracellular fluid, and that the main anion is the chloride ion. In other words the ionic composition of extracellular fluid may be regarded as being built up around sodium chloride, and we may take it that a normal concentration of Na^+ in the plasma (138 to 142 m. equiv/l.) implies normal osmolarity of the extracellular fluid.

The chief cation inside the cells is potassium, whose concentration in intracellular fluid may be regarded as normally roughly equal to the concentration of Na^+ in extracellular fluid. In disease this relationship can become grossly altered. Thus, when the kidney fails to eliminate acid, K^+ is liberated from the cells into the plasma where it may reach a lethal concentration. Conversely, when there is excessive secretion of salt-retaining hormone (p. 938) large amounts of Na^+ may enter the cells, displacing K^+ from them. The cell membranes are freely permeable to water, urea, glucose and electrolytes; it is therefore remarkable that there should be such large differences in the concentrations of Na^+ and K^+ on opposite sides of the cell membranes. The reasons for this are not properly understood. Intracellular fluid contains a relatively high concentration of Mg^{++} and it also contains Na^+ at a concentration approximately one-tenth that found in extracellular fluid. When one considers the relative volumes of intra- and extracellular fluids (Table 35, 3) the amount of sodium within the cells is impressive. The principal anions inside the cell are protein and phosphate, the latter existing chiefly in organic molecules such as adenosine triphosphate, diphosphoglycerate and creatine phosphate; the concentration of inorganic phosphate inside the cell is similar to that outside (2 to 3 m. equiv/l.).

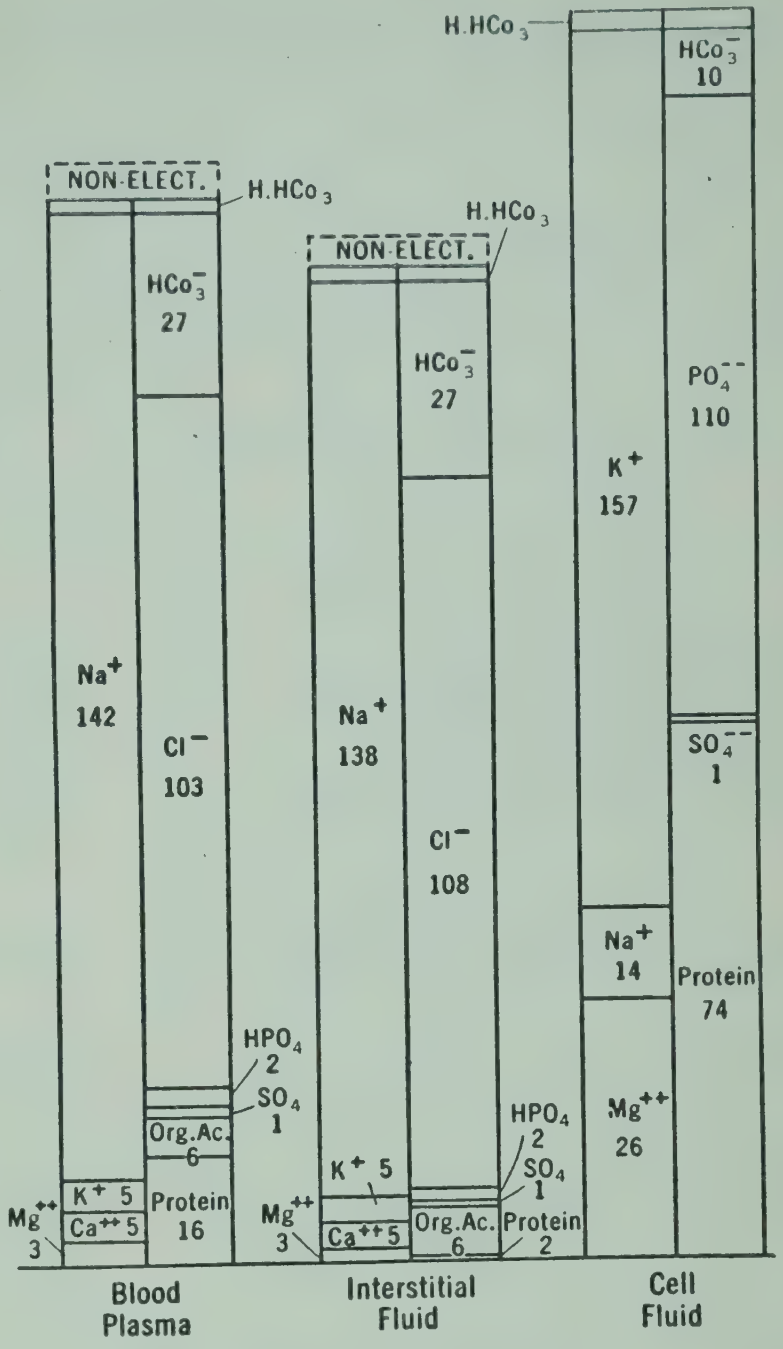


FIG. 35, 5

Electrolyte composition of plasma, interstitial fluid, and intracellular fluid. The principal cation of the extracellular fluid (plasma plus interstitial fluid) is sodium, and the principal anions are bicarbonate and chloride. The principal cations of cell fluid are potassium and magnesium, and the principal anions are phosphate and protein. (Modified from J. L. Gamble (1951) *Companionship of Water and Electrolytes in the Organization of Body Fluids.* Lane Medical Lectures. Stanford Univ. Press.)

Exchangeable sodium and potassium.—The total amount of sodium (or potassium) in the body may be regarded as a space or pool through which any sodium (or potassium) added to the body will become distributed and by which any deficit may be borne. When a known amount of a radioactive isotope of sodium, ^{24}Na, is injected into the blood it at once begins to exchange with the body pool of sodium so that equilibrium with the sodium of the extracellular fluid is reached in about an hour, but with intracellular sodium, especially that in bone, not for some twenty hours. If after that time a specimen of blood is removed and its radioactivity measured, then, after an allowance is made for isotope lost in the urine, the total amount of sodium in which it has been diluted (with which it has exchanged) can be measured. This is the total body sodium, the sodium space, or the exchangeable sodium. As we have seen, this is partly in the cells and partly outside them (Table 35, 6).

TABLE 35, 6

Body Composition as Measured by Dilution in a Normal Man *

Body constituent and method used	Dilution volume or mass (litres or m.equiv)	Per cent. body weight	By surface area (per sq.m)
A—Direct Measurements			
Plasma volume (blue dye)	3·67 litres	5·1	1·92 litres
Extracellular volume (thiocyanate)	16·8 ,,	23·5	8·80 ,,
,, ,, (thiosulphate)	13·4 ,,	18·8	7·03 ,,
Total body water (deuterium)	40·6 ,,	57·0	21·3 ,,
,, exchangeable sodium (^{24}Na)	3075 m.equiv	43·2 m.equiv/kg.	1610 m.equiv
,, ,, potassium (^{42}K)	3440 ,,	48·2 ,, ,,	1801 ,,
B—Derived Data			
‘Intracellular’ water (by difference) †	27·2 litres	38·1	14·2 litres
Extracellular potassium (4·3 m.equiv/l. × thiosulphate space)	57·6 m.equiv	0·81 m.equiv/kg.	..
‘Intracellular’ potassium (by difference)’ †	3382 ,,	47·4 ,, ,,	..
‘Av intracellular’ potassium concentration †	124 m.equiv/litre		
Extracellular sodium (142 m.equiv/l. × thiosulphate space)	1903 m.equiv	26·7 ,, ,,	..
‘Intracellular’ sodium (by difference) †	1172 ,,	16·4 ,, ,,	..
‘Av intracellular’ sodium concentration ‡	43·2 m.equiv/litre		

* Subject: D. C.; male; age, 32; weight, 71·4 kg.; height, 181·5 cm; surface area, 1·91 sq. m.

† Intracellular water and intracellular potassium concentration are dependent on the characteristics of the methods used for estimating extracellular fluid volumes and do not represent exact values in any single tissue. It is particularly important to emphasize that the true intracellular water is not obtained merely by subtracting an extracellular dilution volume from total body water. The transcellular water—at present unknown—must also be subtracted.

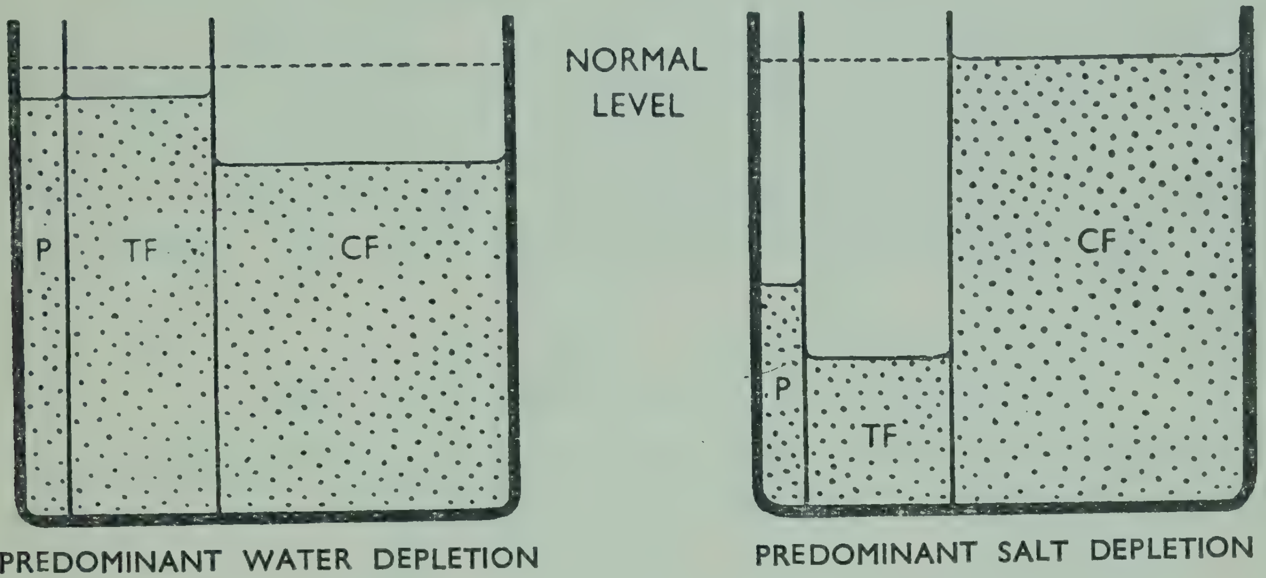
‡ This figure for ‘average intracellular’ sodium includes some bone matrix sodium and is an average of all the sodium not in solution in extracellular fluid. It should not be construed as indicating the exact sodium concentration in, for example, muscle or liver cells.

I. S. Edelman, J. M. Olney, A. H. James, L. Brooks & F. D. Moore (1952), *Science*, **115**, 447.

Potassium exchanges much more slowly and equilibrium is reached at different times in different tissues, perhaps in relation to their metabolic activities. From what has already been said it is obvious that the potassium ‘space’ is almost entirely intracellular (Table 35, 6).

The reason why Na^+ is the principal cation of the extracellular fluid while K^+ is the main cation in the cells is not at present clear. The explanation that the difference is the result of the peculiar properties of the boundary surface, the cell membranes, is no longer tenable since it has been demonstrated that in certain circumstances both Na^+ and K^+ pass across the membrane quite readily. It has been suggested that Na^+ is in fact free to diffuse into the cell but that the differential ionic distribution is the result of a 'sodium pump', that is to say of the expenditure of energy, possibly derived from an oxidation-reduction system. The transport of oxygen across isolated frog skin depends on the oxygen supplied to the skin. Since the oxygen consumption is some 20 to 25 per cent. of the sodium transported, it may be concluded that for each additional atom of oxygen consumed, eight to ten atoms of sodium can be transported. This relatively low oxygen consumption suggests that more than a single step is involved in the oxidative metabolic cycle.

Dehydration.—This term has been applied to two different conditions, both of which have been produced experimentally in man. Though often associated, in their pure forms they can be distinguished clinically. Appropriate treatment depends on understanding the mechanism of their production.



A B

Fig. 35, 7

Diagrams to show the alterations in distribution of the water of the body in predominant water depletion and predominant salt depletion. Compare with Fig. 35, 4. (*After* H. L. Marriott (1947), *Brit. med. J.* i, 245.)

(a) *Predominant water depletion* is due to drastic limitation of water intake, as for example following surgical operations on the alimentary tract. Even if no water is drunk the 'unavoidable loss' of water from the lungs and skin and in the urine and fæces continues, though without a corresponding loss of salt. The extracellular fluid, therefore, tends to become hypertonic, water is withdrawn from the cells and the volume of the extracellular fluid, as evidenced by normal values for the plasma protein concentrations and hæmatocrit, is almost unchanged (Fig. 35, 7A).

(b) *Predominant salt depletion* is the result of loss of Na^+ and Cl^- from the extracellular space in excess of water loss. This occurs for example after excessive sweating if the fluid drunk contains no salt; it may also develop in patients who lose large amounts of gastric juice, the volume of which is replaced by fluid given intravenously containing little or no sodium chloride. The extracellular fluid becomes hypotonic and water passes into the cells to reduce the osmolar concentration of the intracellular fluid (Fig. 35, 7B). The tissue cells swell. The concentration of plasma protein and the hæmatocrit increase and the blood becomes more viscous. This state of affairs is rapidly followed by circulatory failure.

The case of shipwreck survivors adrift in a lifeboat with restricted water and food supplies is of practical importance and serves to illustrate the principles just discussed. The recommendations of a Medical Research Council Committee (1943) may be summarized as follows. Water is much more important than food but, unless there has been much loss of water by sweating, none should be issued in the first twenty-four hours; thereafter 18 oz (0·5 litre) are issued per man per day. This is not sufficient to meet the water requirements but a reasonable state of fitness can be maintained on this ration; smaller intakes lead to rapid deterioration. Although rinsing out the mouth with sea-water may be helpful, sea-water should never be drunk; a large intake of hypertonic sea-water is fatal and small amounts are harmful. The lifeboat rations should consist mainly of fats, starches and sugars. Since salt and the end-products of nitrogen metabolism require water for their excretion the salt and protein content of the diet should be low. A fat diet is valuable as it reduces hunger for a longer time than any other foodstuff (p. 234). Usually defæcation occurs infrequently and the volume of the urine is also reduced; in these ways body water is conserved. The drinking of urine must be prohibited; this fluid is hypertonic and in any case contains substances which may be toxic.

Formation of Interstitial Fluid

The passage of water and dissolved substances through the walls of the capillaries into the tissue fluid depends partly on filtration pressure, that is the hydrostatic pressure within the capillaries, and partly on differences of osmotic pressure between the blood in the capillaries and the tissue fluid outside them. In the capillaries, the non-dialysable constituents, chiefly the plasma proteins, exert a small but highly significant osmotic pressure of about 28 mm Hg. Although this colloid osmotic pressure is small in comparison with the total osmotic pressure of the blood plasma (about 6·5 atmospheres), it is an exceedingly important factor in fluid exchange, because it is exerted on one side only of the capillary membrane.

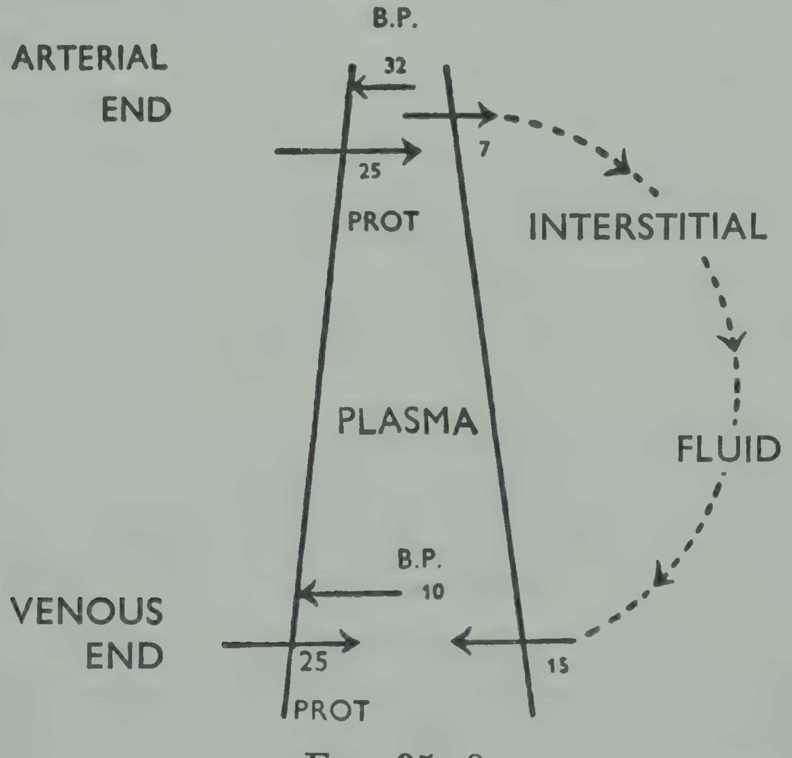


Fig. 35, 8

Capillary vessel showing osmotic and hydrostatic pressures with resultant flow of tissue fluid. B.P., blood pressure. PROT, colloid osmotic pressure of 25 mm Hg. At the arterial end of the capillary the B.P. is 7 mm Hg greater than the colloid osmotic pressure, whereas at the venous end the B.P. is 15 mm Hg less than the colloid osmotic pressure.

Landis's measurements of capillary blood pressure (p. 490) help us to form an estimate of the forces governing the passage of fluid across the capillary endothelium (Fig. 35, 8). The pressure at the arterial end of the capillary is such as to increase the tissue fluid at the expense of the plasma, whereas at the venous end of the capillary fluid tends to pass in the reverse direction. The circulation of tissue fluid in the direction of the arrows (Fig. 35, 8), brought about in this way, facilitates the exchange of food materials, metabolites and salts, and helps to maintain the constancy of the environment of the tissue cells. The relation between the capillary blood pressure, the colloid osmotic pressure of the blood and the movement of fluid across the capillary walls has been beautifully demonstrated by Landis in the frog (Fig. 35, 9). When the capillary pressure equals the colloid osmotic pressure no fluid movement occurs. At capillary pressures in excess of this (about 11·5 cm of water in the frog) filtration is predominant, while at lower pressures fluid is absorbed into the blood.

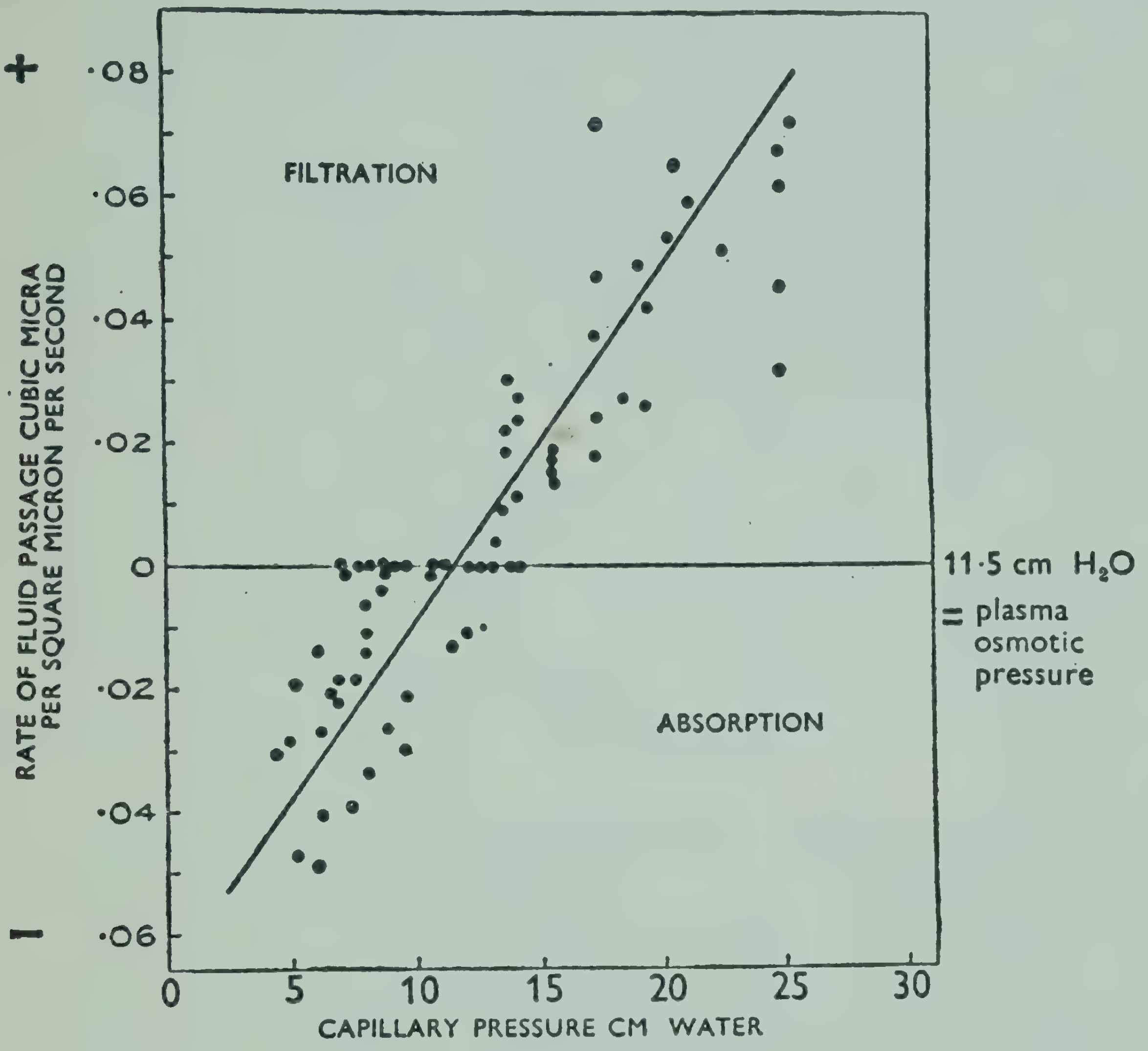


FIG. 35, 9

The effect of pressure on the passage of fluid through the capillary wall. By means of a micro-manipulator a mesenteric vessel of the frog was blocked by pressure with a blunt-ended glass rod. The red cells moved away from the block if absorption took place. The pressure in the capillary was measured by a fine pipette attached to a water manometer. Positive values of the rate of fluid passage indicate filtration of fluid; negative values indicate absorption (E. M. Landis (1927), *Amer. J. Physiol.* **82**, 217).

In man injection of radioactive sodium (^{24}Na) intravenously shows that about 32 per cent. of the total plasma sodium diffuses out of the vascular bed per minute into the tissue fluid; this corresponds to 4 kg. of sodium per day for an average man. The same amount, of course, diffuses back into the blood stream. The amount of water passing in and out is even greater —about 6000 litres per day.

The speed of the movement of the electrolytes in and out of the capillaries is very rapid. If ^{24}Na is given intravenously to a man who is sweating profusely the radiosodium can be detected in the sweat in 75 seconds. Diffusion into bone, vitreous and aqueous humours and cerebrospinal fluid is much slower.

When a muscle contracts or a gland secretes the metabolic processes involve the breakdown of large molecules, each giving rise to numerous smaller ones. This increased concentration of molecules raises the osmotic pressure of the intracellular fluid by as much as 50 per cent., and water enters the cells from the extracellular space. An idea of the volume involved can be obtained if we know both the blood flow through the muscle and the increase in hæmoglobin concentration as the blood passes from the arteries to the veins. In the resting muscles of man there is a negligible amount of tissue fluid, but in exercise as much as 2 ml. per 100 g. of muscle may be produced per minute. This rapid passage of fluid from the capillaries is facilitated by the increase of pressure at the arterial end of the capillaries resulting from dilatation of arterioles and by the increased endothelial surface made available for filtration by the opening up of many capillaries normally closed (p. 518). It is not so easy to see how the fluid is removed after the period of activity. Presumably when the capillary pressure is lowered during rest the metabolites can diffuse back into the capillaries. Much of the fluid is absorbed into the lymphatic channels for the flow of lymph from a contracting muscle or a secreting gland is greatly increased over the resting values. The factors governing the entrance of fluid into the lymphatic system are, however, unknown.

The Lymphatics.—The lymphatics begin in the periphery, as a closed system of tubes consisting of an endothelial lining supported by fibrous tissue. The larger vessels have in addition muscle fibres in their walls. The lymphatic vessels are well provided with valves and the flow of lymph in them is brought about by muscular and respiratory movements in much the same way as the flow of blood in the veins. Although the composition of lymph varies very much, according to the area from which it comes and the activity of the part, the protein content is intermediate between that of tissue fluid and blood plasma. No satisfactory explanation for this relatively high protein concentration has yet been offered.

Lymph Nodes.—Before reaching the blood the lymph passes through one or more lymph nodes (Fig. 35, 10). The lymph enters a node by several afferent lymphatic vessels and flows into the *subcapsular lymph space* (peripheral sinus), which is lined by macrophages. Back flow is prevented by a valve at the point of entry of each afferent vessel. The lymph spreads along the subcapsular space and is guided by the trabecular framework through the *cortex* of the node, which is densely packed with lymphocytes, to the more loosely packed *medulla*, from which it escapes by the single efferent lymphatic vessel. In the dense cortical area are paler germ centres containing developing lymphocytes which do not stain deeply with basic dyes. The majority of the lymphocytes are formed in these centres.

The lymph from all over the body is collected into larger and larger lymphatic vessels formed by the union of smaller vessels. From most of the body, and the whole of the intestine, the lymph passes to the *thoracic* *duct* which enters the large veins at the root of the neck on the left side. The lymph from the right upper half of the body enters the great veins via the *right lymphatic duct*. Back flow of venous blood into the lymphatic system is prevented by valves placed at the orifices of the ducts. By subcutaneous injection of dyes dissolved in saline solution it can be shown that the flow of tissue fluid along the lymphatic vessels is extremely slow, although the dyestuffs are quickly absorbed into the capillary blood vessels and are rapidly excreted in the urine. Particulate matter, such as bacteria or India ink, is carried in the lymphatics to the lymph nodes which act as filters preventing the entry of dye or bacteria into the general circulation.

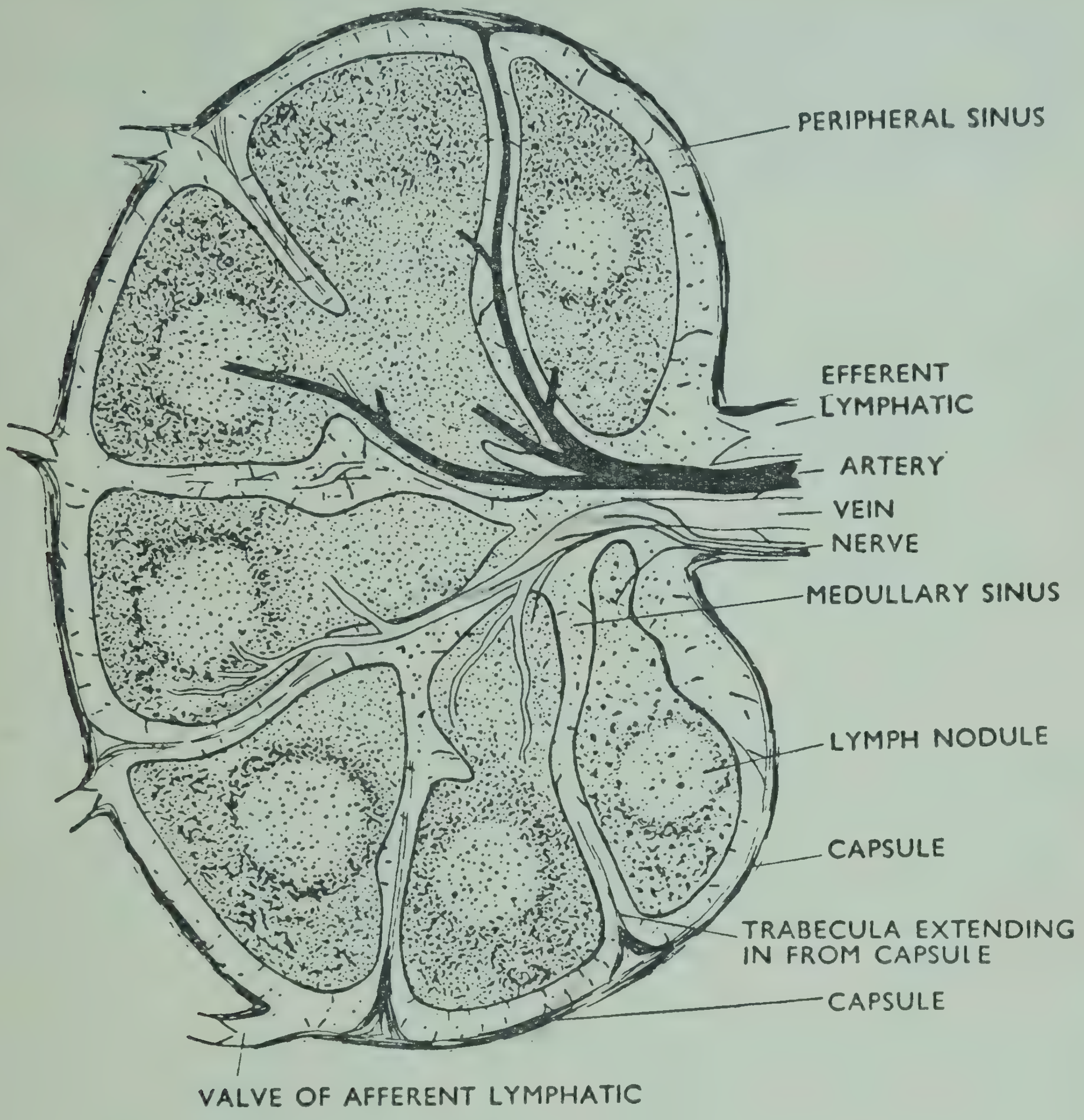

PERIPHERAL SINUS
EFFERENT LYMPHATIC
ARTERY
VEIN
NERVE
MEDULLARY SINUS
LYMPH NODULE
CAPSULE
TRABECULA EXTENDING IN FROM CAPSULE
CAPSULE
VALVE OF AFFERENT LYMPHATIC

Fig. 35, 10

Diagram of a typical lymph node. The lymph enters the peripheral sinus by many afferent lymphatics guarded by valves. It percolates in the medullary sinuses between the masses of lymphoid tissue and leaves at the hilus by a large efferent vessel. Arteries (black), veins (clear) and nerves enter at the hilus. The pale areas in the middle of the lymph nodules are the germ centres. × about 10 (Cowdry, *Textbook of Histology*, 1938, p. 154. London: Kimpton.)

In the fasting state the lymph from the thoracic duct is a pale yellow transparent fluid but after a fatty meal it becomes milky (p. 270), and this milky appearance is even more pronounced in the lymphatics of the small intestine which are often called *lacteals*.

HYDROGEN-ION CONCENTRATION OF THE EXTRACELLULAR FLUID

Variations in the output of ADH control the elimination of water by the kidneys so as to maintain approximately constant the total solute concentration, that is the osmolarity, of the extracellular fluid. Since sodium is the major cation of the extracellular fluid this operation implies that the plasma concentration of sodium shall be held within relatively narrow limits. Even relatively large alterations in the concentrations of K^+, Ca^{++} and Mg^{++} can have little effect on the osmolarity.

Fig. 35, 11 shows that the sum of the cation concentrations (155 m. equiv/litre) is equal to the total concentration of anions, a state of affairs that is essential to the maintenance of electrical neutrality. The normal processes of metabolism result in the production of phosphoric, sulphuric and lactic acids as well as large amounts of CO_2 which combines with H_2O in the extracellular fluid to form H_2CO_3. Being a weak acid this is only partly dissociated thus:

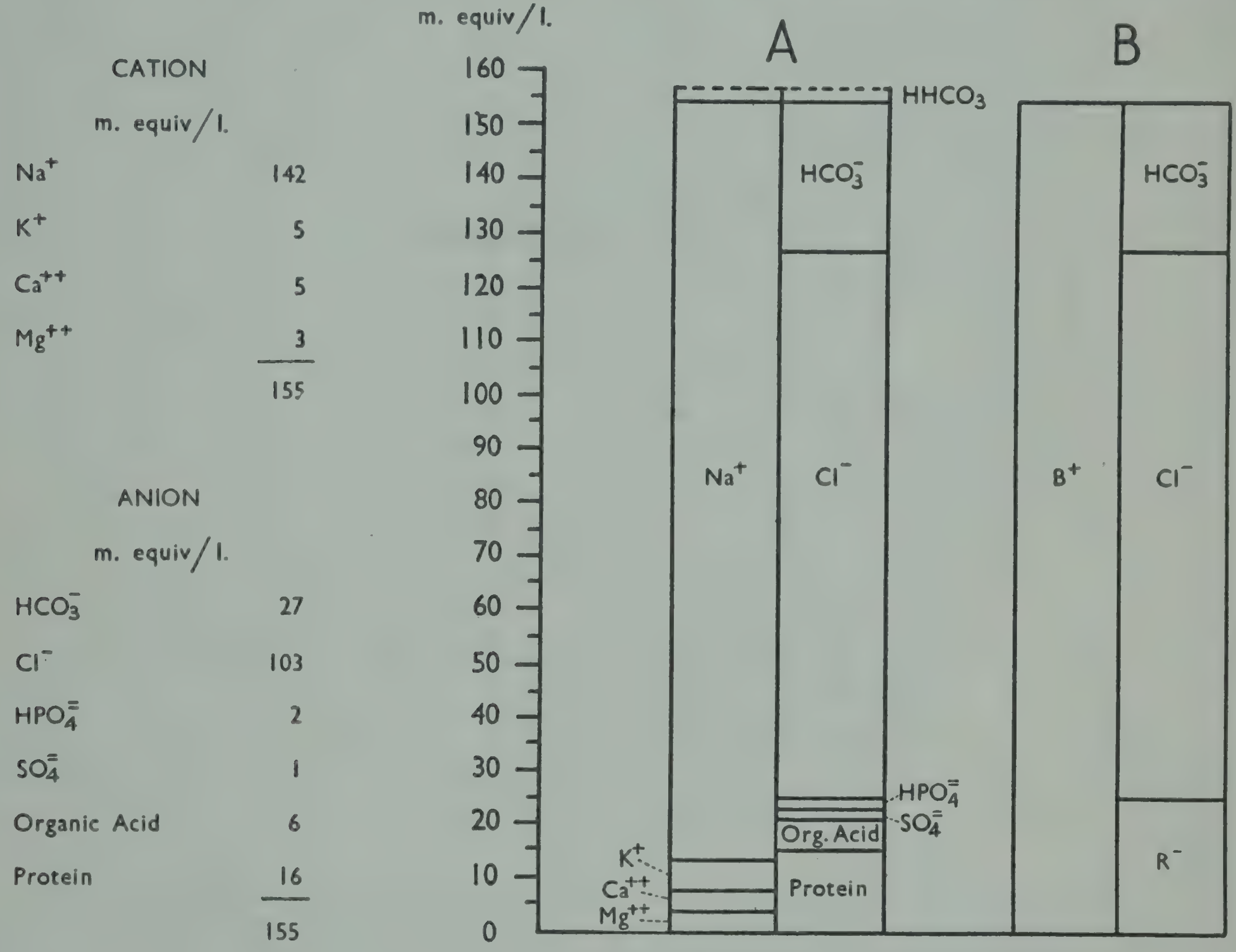


FIG. 35, 11

The ionic composition of normal plasma (after Gamble). The part of the diagram marked A gives the composition in detail; that marked B contains the data usually adequate for clinical purposes. The latter requires the determination only of total cation (or of sodium + 13), of chloride, and of bicarbonate (A. Leaf & L. H. Newburgh (1955), *Significance of the Body Fluids in Clinical Medicine.* 2nd Ed. Springfield: C. C. Thomas.

$$H_2CO_3 \rightleftharpoons H^+ + HCO_3^-$$

The bicarbonate ion HCO_3^- is balanced by cation to the extent, under normal conditions, of 27 m. equiv/litre. The hydrogen ion concentration of the extracellular fluid, that is to say its reaction or pH is thus established by the degree of dissociation of carbonic acid.

If K is the dissociation constant for carbonic acid then

$$K = \frac{[H^+][HCO_3^-]}{[H_2CO_3]}$$

where $[H^+]$ is the concentration of hydrogen ions,
$[HCO_3^-]$,, ,, ,, bicarbonate ions, and
$[H_2CO_3]$,, ,, ,, undissociated carbonic acid,

or

$$[H^+] = K\frac{[H_2CO_3]}{[HCO_3^-]} \quad . \quad . \quad . \quad . \quad (1)$$

Taking the negative logarithm of both sides we have

$$-\log [H^+] = -\log K - \log \frac{[H_2CO_3]}{[HCO_3^-]}$$

$$= -\log K + \log \frac{[HCO_3^-]}{[H_2CO_3]}$$

$$pH = pK + \log \frac{[HCO_3^-]}{[H_2CO_3]} \quad . \quad . \quad . \quad (2)$$

This is known as the Henderson-Hasselbalch equation. pK, which is the pH value when $[HCO_3^-] = [H_2CO_3]$, has the value 6·1.

When equation (2) is applied to blood it is seen that the pH of the blood depends upon the ratio $\frac{[HCO_3^-]}{[H_2CO_3]}$ which is normally about 20. Since log 20 is 1·3, pH = 6·1 + 1·3 = 7·4.

Any circumstance that increases the ratio of carbonic acid to bicarbonate ion leads to an increase in $[H^+]$, that is to say *acidosis*; conversely, anything that diminishes the ratio tends to cause a reduction in $[H^+]$, that is *alkalosis*.

The total carbon dioxide of the blood is the sum of the free carbon dioxide and the combined carbon dioxide. The readily available supply of base, that is to say the base combined with carbonic acid as bicarbonate ($BHCO_3$), was called by Van Slyke the *alkali reserve* to distinguish it from the total base which includes sodium, potassium, magnesium and calcium combined as chlorides and other salts from which the acid is not displaceable by carbon dioxide. Since blood takes up carbon dioxide in proportion to its available base, the alkali reserve can be assessed by measuring what is called the *carbon dioxide combining power*. A sample of venous blood is first exposed to normal alveolar air in a tonometer and then its total carbon dioxide content is determined. This gives a rough measurement of the carbon dioxide combining power and so of the alkali reserve, but to determine the latter accurately it is necessary to subtract the free carbon dioxide in the blood. This value can be obtained by making use of the fact that the carbon dioxide tension in the alveolar air is very nearly the tension in the arterial blood. If the percentage of carbon dioxide in the alveolar air is measured, the free carbon dioxide in the blood can be calculated from its partial pressure and its coefficient of solubility (0·51) (see p. 571). When the free carbon dioxide is deducted from the total carbon dioxide we have the alkali reserve in terms of carbon dioxide combining power. The normal range of alkali reserve is considerable, namely 24 to 34 m. equiv per litre.

A tendency to an alteration in the reaction of the blood is followed by compensatory mechanisms which operate to maintain the pH within the normal range of 7·3 to 7·5. Should the adjustment be inadequate then the pH will exceed these limits and either severe acidosis or alkalosis will result. The compensatory mechanisms are of two kinds, respiratory and renal. Since the respiratory centre and the chemoreceptors in the aortic and carotid bodies are normally very sensitive to changes of pH a slight increase in the hydrogen-ion concentration of the blood increases the respiratory minute volume and more carbon dioxide is eliminated. If the depth of respiration becomes unduly great the patient is described as having 'air hunger'. Conversely a reduction in hydrogen-ion concentration is associated with slow shallow breathing, the effect of which is to retain carbon dioxide in the body.

Alterations in the rate and depth of respiration, which can be effected rapidly, do, therefore, tend to prevent alterations in the pH of extracellular fluid. However, the respiratory apparatus has no power to remove from the blood any excess of acid or alkali. The removal of such abnormalities is the work of the kidneys and is accomplished more slowly.

The kidneys can produce considerable alterations in the pH of the urine without damage to themselves. They accomplish such changes in a number of ways.

(*a*) By a process of cation exchange in the renal tubules as the result of which H^+ ions are exchanged for fixed cations (Na^+, K^+, Ca^{++}, and Mg^{++}) in the tubular urine. The H^+ ion is eliminated in the urine and the fixed cations conserved to the body.

(*b*) When the rate of formation of acid in the body is increased, as for example in diabetic ketosis (p. 330), or when certain salts, such as ammonium chloride, are ingested, the kidney may excrete the anion of the acid. Equivalent amounts of the fixed cations (Na^+, K^+, Ca^{++} or Mg^{++}) of the extracellular fluid have, however, to be excreted simultaneously.

(*c*) The kidney can synthesize ammonia (p. 621, which, liberated into the acid tubular urine, captures H^+ ions forming the weakly basic cation NH_4^+. The acidity of the urine is thus reduced.

(*d*) The renal tubules can vary the rate of reabsorption of the large amounts of HCO_3^- presented to them in the glomerular filtrate. By reduction of the rate of reabsorption an excess of alkali (bicarbonate) can be eliminated. The tubular reabsorption of HCO_3^- is regulated by the concentration of H_2CO_3 in the extracellular fluid. An increase in carbonic acid concentration increases the reabsorption of HCO_3^- and a decrease depresses it.

Acidosis and Alkalosis

We have seen that the pH of the extracellular fluid is determined by the ratio of the concentrations of bicarbonate ion and carbonic acid (equation 2, p. 641). Alterations in the concentrations of either of these may therefore disturb pH.

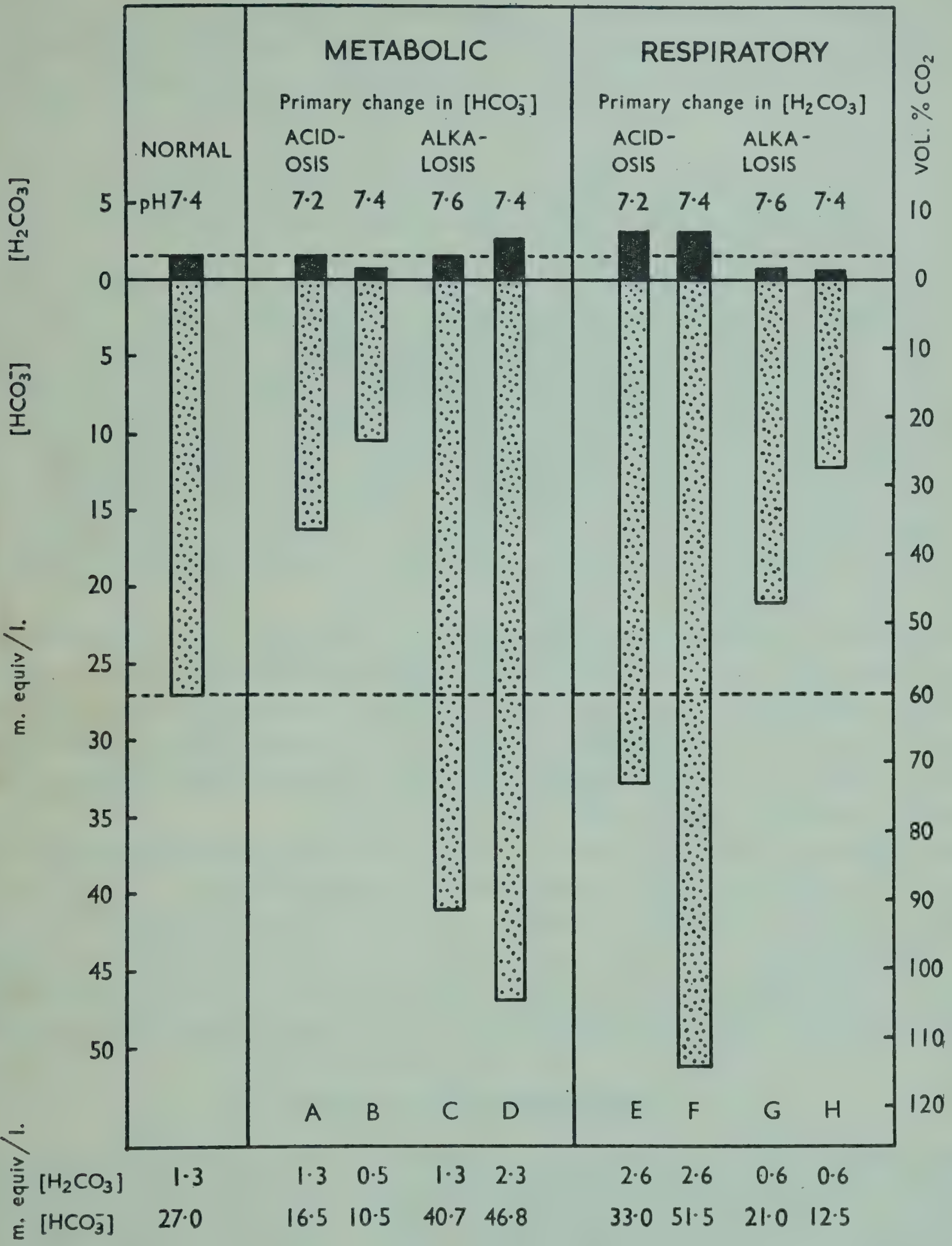


Fig. 35, 12

To illustrate the main varieties of acidosis and alkalosis. (*After* J. L. Gamble (1954) *Chemical Anatomy, Physiology and Pathology of Extracellular Fluid.* 6th Ed. Cambridge, Mass.: Harvard Univ. Press.

Metabolic Acidosis and Alkalosis.—There are a number of metabolic disturbances that give rise to primary alterations in [HCO_3^-]. Such alterations are accompanied by secondary changes in [H_2CO_3], the effect of which is to restore the value of the ratio and thus preserve the normal pH.

Disturbances of pH that occur because of lack of availability of base or from accumulation of acid other than carbonic acid are said to be metabolic.

Metabolic acidosis may arise because of excessive production of acid, for example of lactic acid in severe exercise, or of aceto-acetic and β-hydroxybutyric acids in diabetic ketosis (p. 330). It may also follow the ingestion of large amounts of ammonium chloride. The loss from the body of abnormal amounts of cation, such as happens for example with severe diarrhœa unaccompanied by vomiting, may result in metabolic acidosis, as can also a breakdown of the renal mechanisms for the elimination of H^+ ion.

Metabolic alkalosis, on the other hand, may result from excessive loss of anion other than carbonic acid, for example Cl^- in recurrent vomiting; or from the ingestion of unusually large amounts of alkali.

The alterations in $[HCO_3^-]$, that is bicarbonate ion, and in $[H_2CO_3]$, carbonic acid, in metabolic acidosis and alkalosis are shown diagrammatically in Fig. 35, 12. Actual values for the concentrations are given at the foot of the diagram, from which it can be seen that determination of $[B.HCO_3]$, the CO_2-combining power gives a reliable index of the situation.

Respiratory Acidosis and Alkalosis.—Here the alterations in pH are determined primarily by changes in $[H_2CO_3]$, that is to say by changes in pulmonary ventilation. Subsequent alterations in $[HCO_3^-]$ occur which have the effect of restoring the value of the ratio to normal and so preserving a normal pH.

Respiratory acidosis is usually due to defective elimination of carbon dioxide from the lungs, either because of chronic disease of the lungs which interferes with the exchange of carbon dioxide between blood and alveolar air, or because the respiratory centre has become relatively insensitive to concentrations in the blood of carbon dioxide that would otherwise stimulate increased ventilation.

Respiratory alkalosis results from excessive elimination of carbon dioxide through the lungs as the result of hyperpnœa.

The values for $[HCO_3^-]$ and $[H_2CO_3]$ at the foot of Fig. 35, 12 show that determinations of CO_2-combining power do not indicate respiratory acidosis or alkalosis. Such abnormalities can be detected only by determination of both $[H_2CO_3]$ and $[HCO_3^-]$ or by measuring the blood pH.

CEREBROSPINAL FLUID

The central nervous system is covered by three meninges (Fig. 35, 13). Inside the tough outer coat called the *dura mater* is a narrow space containing a very small amount of tissue fluid. This is bounded internally by the *arachnoid mater* which is separated from the *pia mater* by the *subarachnoid space* containing *cerebrospinal fluid*. The subarachnoid space does not communicate with the subdural space, but is continuous with the ventricular system of the brain through medial and lateral apertures in the roof of the fourth ventricle. A sample of cerebrospinal fluid can be obtained in man by passing a long needle between the occipital bone and the atlas into the cisterna cerebellomedullaris at the base of the brain (*cisternal puncture*) (Fig. 35, 16) or between the third and fourth lumbar spinous processes into the spinal subarachnoid space (*lumbar puncture*) (Fig. 35, 17). The latter is the safer procedure because, although the dura and the arachnoid meninges

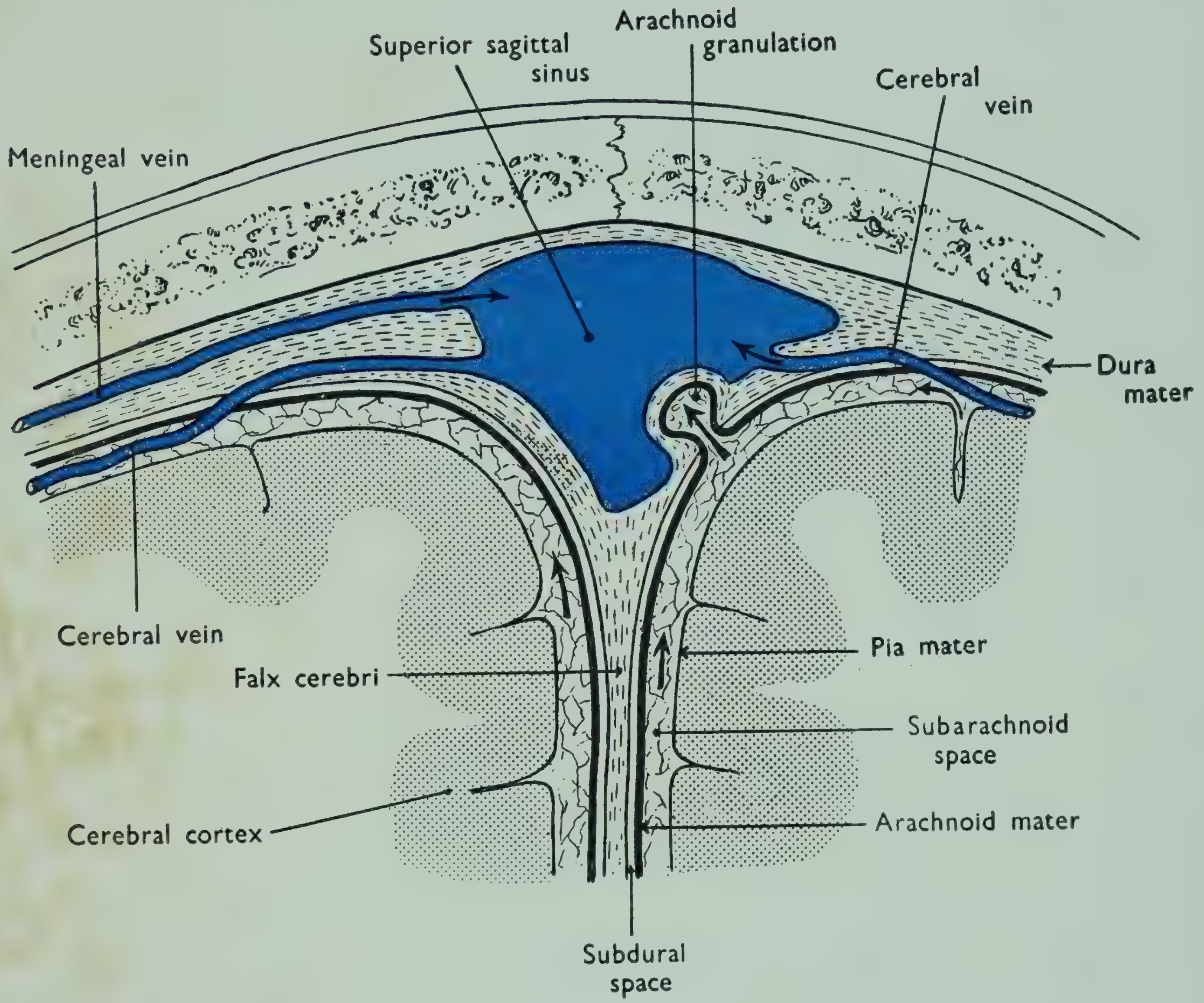

Arachnoid granulation
Superior sagittal sinus
Cerebral vein
Meningeal vein
Dura mater
Cerebral vein
Pia mater
Falx cerebri
Subarachnoid space
Cerebral cortex
Arachnoid mater
Subdural space

FIG. 35, 13

Diagram of the membranes of the brain to show the venous drainage and the route of escape of the cerebrospinal fluid.

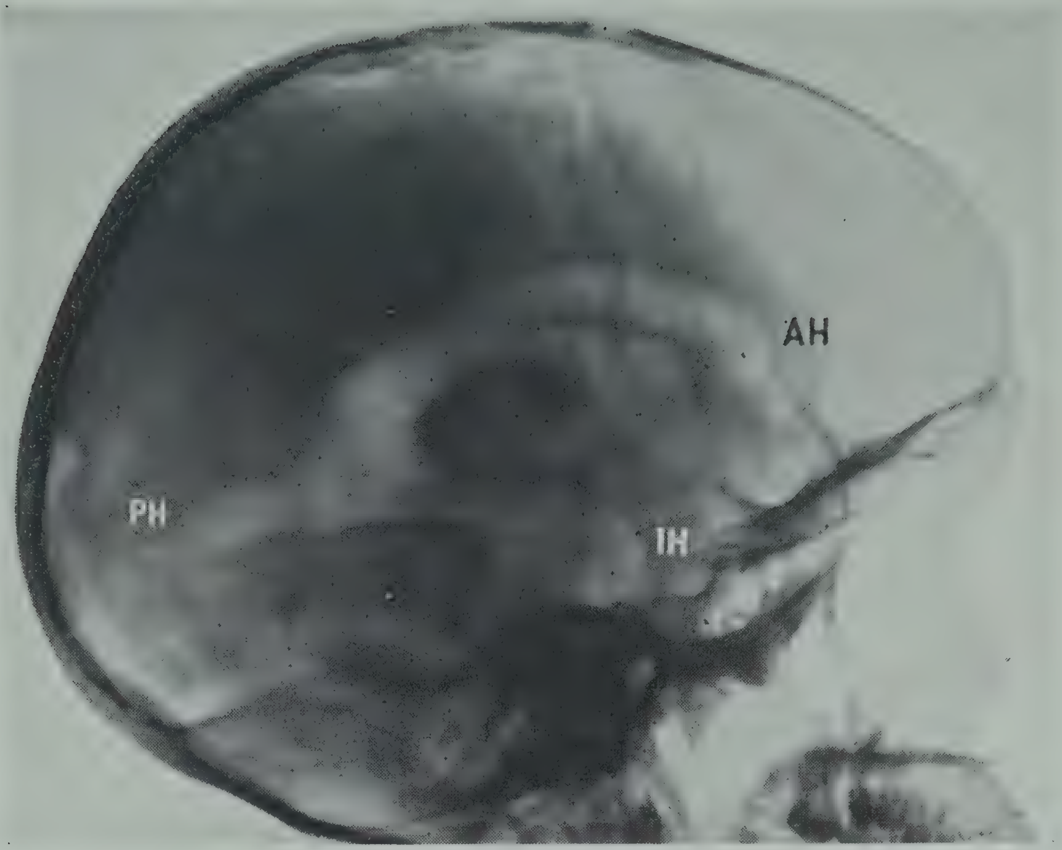


Fig. 35, 14

Ventriculogram (lateral view). The ventricular cavities, being filled with air, show as lighter areas. AH, anterior horn. PH, posterior horn. IH, inferior horn.

(*By courtesy of J. Sloan Robertson.*)

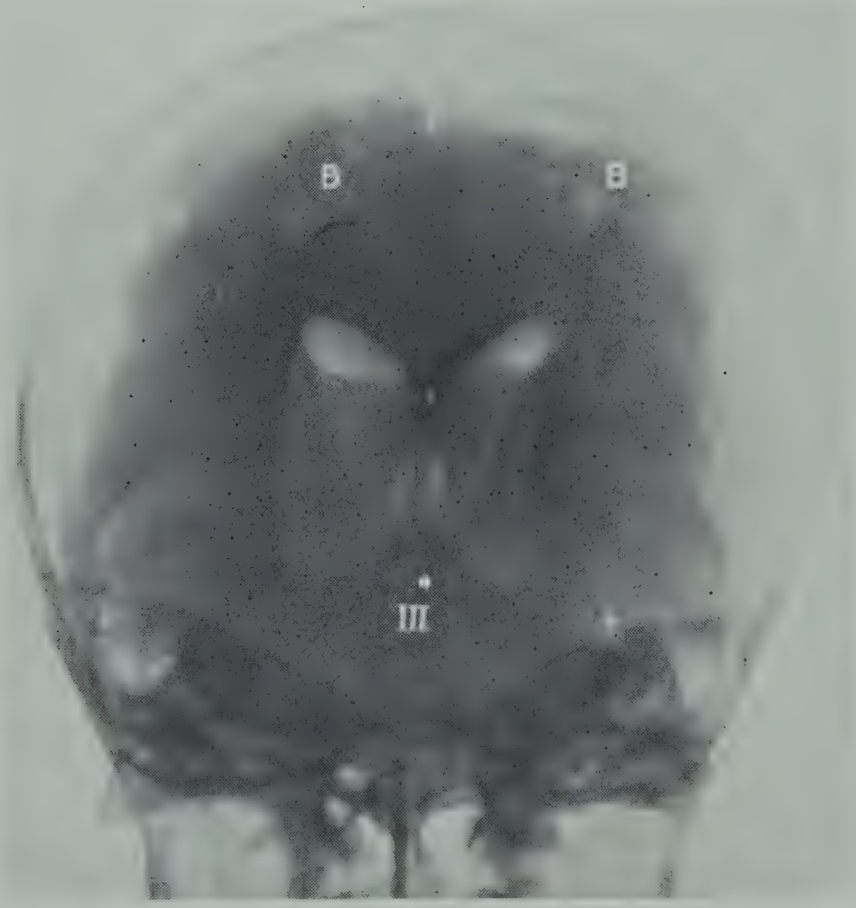


Fig. 35, 15

Ventriculogram (antero - posterior view). The anterior horns can be recognized above the inferior horns. The third ventricle is above the point marked III. B, burr hole.

(*By courtesy of J. Sloan Robertson.*)

end at the second sacral vertebra, the cord comes to an end at the level of the first lumbar vertebra and so cannot be damaged by the needle which enters the subarachnoid space an inch or so lower.

Normal cerebrospinal fluid (Table 35, 18) is a clear colourless fluid of pH 7·4, and sp. gr. 1005, containing not more than 5 lymphocytes per cu.mm. Its composition is similar to that of plasma except that it contains very much less protein (only 0·02 per cent.). The total volume is about 120 ml. If a vertical tube is attached to the lumbar puncture needle the fluid rises about 10 cm with the patient lying on his side and about 30 cm with the patient sitting. A rise of intracranial venous pressure produced by coughing or by digital compression of the internal jugular veins causes the cerebrospinal-fluid pressure to rise rapidly above these values (*Queckenstedt's sign*) and there is a rapid fall in the pressure when the compression is removed.

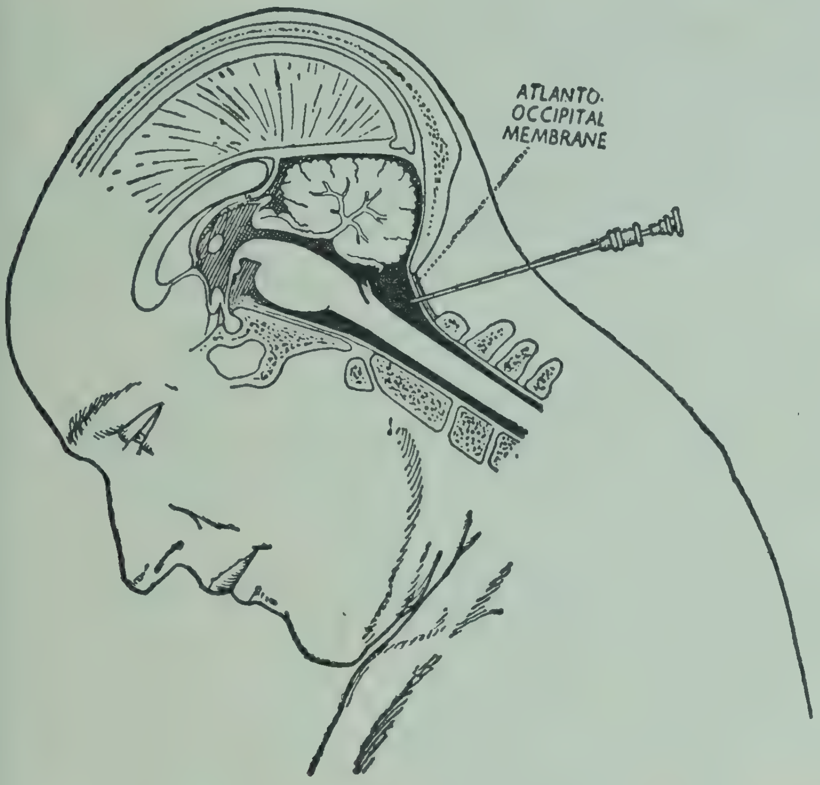


FIG. 35, 16

Cisternal puncture. The diagram shows the direction of the needle (Hamilton, Bailey (1947), *Pye's Surgical Handicraft*, 15th ed., Fig. 520, p. 400. Bristol: Wright).

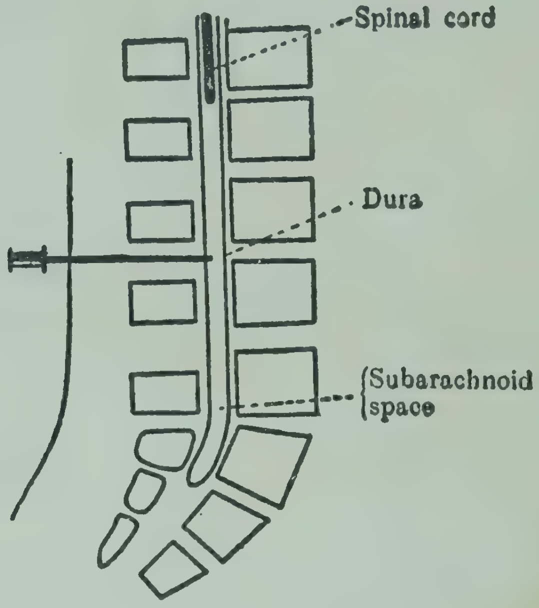


FIG. 35, 17

Route of needle in performing lumbar puncture. The cord is out of danger. The cauda equina is not shown (A. Lee McGregor (1946), *Synopsis of Surgical Anatomy*, 6th ed., Fig. 626, p. 695. Bristol: Wright).

The very vascular *choroid plexuses* which project into all the four ventricles of the brain are the source of the cerebrospinal fluid. They are fringe-like processes of pia mater, covered by a layer of cells continuous with the ciliated epithelium lining the ventricles (ependyma). The cells of the choroid plexuses seem to exert a selective action on the substances passing through them, because the composition of plasma and fluid is not quite identical and the distribution of ions between them is not that expected from the Donnan equilibrium. The formation of cerebrospinal fluid from blood plasma is a process of secretion rather than of simple diffusion or filtration (Table 35, 18). The cells of the plexuses are sometimes described as the *blood-brain barrier* because certain substances, for example sodium iodide and sodium nitrite, are held back and do not pass into the cerebrospinal fluid from the blood; even in deep jaundice bile pigment does not enter the fluid. Radioactive ions injected intravenously accumulate more slowly in the cerebrospinal fluid than in the tissue fluid; the rate of entry of potassium ions is much greater than that of phosphate or iodide ions.

The main sources of cerebrospinal fluid are the choroid plexuses in the lateral ventricles, from which the fluid passes into the third ventricle by the interventricular foramina and then by the aqueduct of the mid-brain to the fourth ventricle, from which it escapes by three foramina into the subarachnoid space round the brain and spinal cord. That this is the manner of flow is shown by the results of blockage of any of the foramina; the ventricles upstream then dilate, giving the condition known as *internal hydrocephalus.* Blockage of the cerebral aqueduct, for example, causes distension of the third ventricle and the two lateral ventricles unless the choroid plexuses have previously been removed. These experiments show that there is a continuous production of fluid and, since the pressure of the cerebrospinal fluid remains relatively constant, production must be balanced by an equal absorption of fluid. There is no lymphatic system in the brain and spinal cord. Such of the extracellular fluid as is not reabsorbed back into the vessels through the capillary walls flows into the subarachnoid space along minute prolongations of this space which follow the blood vessels. This process, analogous to the formation of lymph, provides an alternative source for cerebrospinal fluid. By injecting a dye into the subarachnoid space it can be shown that most of the fluid passes through the arachnoid villi which project into the venous sinuses of the dura (Fig. 35, 13) and so enters the blood. If the villi are blocked by injecting lampblack into the subarachnoid space chronic *external hydrocephalus* results. The fluid in the spinal theca is absorbed into the spinal veins.

TABLE 35, 18

Comparison of Chemical Composition of Blood Plasma and Cerebrospinal Fluid

	Blood Plasma mg./100 ml.	Cerebrospinal Fluid mg./100 ml.
Protein	7000–8000	20–30
Urea	16–35	10–40
Glucose	68–96	50–80
Sodium	318–342	350
Potassium	15–20	15
Calcium	10–11	5
Chloride (as Cl^-)	360–378	437–455
Bicarbonate (vol. CO_2 per 100 ml.)	50–75	50–75
Inorganic phosphate	2–5	1–2

The calcium content of the cerebrospinal fluid is relatively low but only half of the calcium of the blood plasma is ionized, so that the two fluids can be considered to be in equilibrium in respect of calcium. Note the higher chloride and lower glucose content of cerebrospinal fluid as compared with blood plasma.

The absorption of cerebrospinal fluid depends on the physical factors discussed already in connexion with the movement of tissue fluid. The hydrostatic pressure of the cerebrospinal fluid is greater than the venous pressure, while the colloid osmotic pressure of the blood plasma, because of its proteins, is greater than that of the cerebrospinal fluid. These differences both favour absorption of water into the veins. The veins do not collapse under the greater cerebrospinal fluid pressure because they are held open by the fibrous dura, as shown in Fig. 35, 13.

Cerebral tumours often interfere with the circulation of the cerebrospinal fluid and a rise of intracranial pressure results. Since the optic nerve and the retina are really parts of the brain this rise of pressure causes a distension and tortuosity of the veins of the retina, with a swelling of the optic disc (*papillœdema*) easily seen with an ophthalmoscope. Increased intracranial pressure may be temporarily relieved by the intravenous injection of hypertonic solutions, such as 30 or 50 per cent. NaCl, or 50 per cent. sucrose. The resultant increase in the osmotic pressure of the blood drains fluid from the brain into the cerebral capillaries, and produces a shrinkage of the brain and a reduction of intracranial pressure which lasts for several hours. The increased osmotic pressure in the vessels of the choroid plexuses may be great enough to reverse the direction of flow so that cerebrospinal fluid, and dye carried in it, may actually be absorbed through the plexuses into the blood stream.

The shape, size and position of the cerebral ventricles often give valuable clues to the position of tumours in the cranial cavity. To make a *ventriculogram* a needle passed through a trephine hole in the skull is pushed into the lateral ventricle; the cerebrospinal fluid drains off and is replaced by air. An X-ray photograph (Figs. 35, 14; 35, 15) shows lighter areas (in the positive) corresponding to the cerebral ventricles.

REFERENCES

Bass, D. E. & Henschel, A. (1956). Responses of body fluid compartments to heat and cold. *Physiol Rev.* **36,** 128–144.

Dougherty, T. F. (1952). Effect of hormones on lymphatic tissue. *Physiol. Rev.* **32,** 379–401.

Drinker, S. K. & Yoffey, J. M. (1941). *Lymphatics, Lymph and Lymphoid Tissue.* Harvard University Press.

Elkinton, J. R. & Danowski, T. S. (1955). *The Body Fluids. Basic Physiology and Practical Therapeutics.* London: Baillière, Tindall & Cox.

Flexner, L. B. (1934). The chemistry and nature of the cerebrospinal fluid. *Physiol. Rev.* **14,** 161–187.

Friedemann, U. (1942). Blood-brain barrier. *Physiol. Rev.* **22,** 125–145.

Gamble, J. L. (1954). *Chemical Anatomy, Physiology and Pathology of Extracellular Fluid.* 6th Ed. Cambridge, Mass.: Harvard Univ. Press.

Hardy, J. D. (1954). *Fluid Therapy.* London: Henry Kimpton.

Kinmonth, J. B. (1954). Lymphangiography in clinical surgery. *Ann. Roy. Coll. Surg. Eng.* **15,** 300–315.

Landis, E. M. (1934). Capillary pressure and capillary permeability. *Physiol. Rev.* **14,** 404–481.

Le Quesne, L. P. (1954). Fluid balance in surgical practice. London: Lloyd-Luke.

Lymph: a symposium (1946). *Ann. N.Y. Acad. Sci.* **46,** 679–882.

Manery, J. F. (1954). Water and electrolyte metabolism. *Physiol. Rev.* **34,** 334–417.

Marriott, H. L. (1950). *Water and Salt Depletion.* Oxford: Blackwell.

Newburgh, L. H. & Leaf, A. (1955). *Significance of Body Fluids in Clinical Medicine.* 2nd Ed. Springfield: C. C. Thomas.

Peters, J. P. (1952). The problem of cardiac œdema. *Amer. J. Med.* **12,** 66–76.

Pinson, E. A. (1952). Water exchange and barriers as studied by the use of hydrogen isotopes. *Physiol. Rev.* **32,** 123–134.

Starling, E. H. (1909). *The Fluids of the Body.* London: Constable.

Wilkinson, A. W. (1955). *Body Fluids in Surgery.* Edinburgh: Livingstone.

CHAPTER 36

SPECIAL SENSES

INTRODUCTION

THERE is no direct relationship between a *stimulus* and the *sensation* it evokes although in ordinary speech the two are often taken to be the same. For instance a sound is often described as loud but the word loud refers to the sensation and not to the stimulus which is more correctly described as intense or powerful. A vibrating tuning-fork placed on the lower end of the radius gives rise to a sensation of vibration, whereas the same fork placed on the head or near the ear arouses a sensation of sound. If a variety of stimuli is applied to any one end-organ the same kind of sensation is produced. In the case of the eye, an electric current, pressure and light all evoke sensations of light. Experiments with single nerve fibres show that all nerve impulses, however they are initiated, are the same. In other words, the same stimulus applied to different end-organs evokes different sensations: different kinds of stimuli applied to one end-organ all call forth the same sensation. Müller called this the law of specific irritability or the law of specific nerve energy. A reasonable conclusion is that the sensation aroused by the stimulation of any end-organ or nerve depends on the route taken by the nerve impulses and on their destination in the central nervous system.

Vision, hearing, taste, olfaction, and the cutaneous senses are included in the general title 'special senses'. In investigating them it is necessary to distinguish between *sensation* and *perception*. Sensation is the process of response by the sensory apparatus to a stimulus derived from an external object. Perception is the process of becoming aware of the object originating the stimulus. When that object is further recognized or identified in accord with previous experience, we have *apperception*. It is the perception and the apperception of objects that normally interests us, and nearly everyone who is given a familiar article to feel describes the object rather than the sensation. If I lift an apple the contribution of sensation would be touch, pressure or weight and, say, cold. Perception would be expressed as "In my hand I've got something round, smooth, moderately heavy and cold". Apperception would be "This feels like an apple".

The absence of a direct relationship between stimulus and sensation is a serious handicap in the investigation of the special senses. We are not entitled to say that one light is twice as bright as another because we have no way of measuring the intensity of our sensations. Only two types of experiment are open to us: (1) measurement of the smallest stimulus which elicits a sensation and (2) matching of sensations produced by simultaneous or successive stimuli. The first measurement is very useful, although the threshold varies to some extent with the experimental method and the condition of the subject. In the second, the method of matching, there is always an error because there is a minimum perceptible difference below which we cannot distinguish two stimuli. Weber's law states that this minimum difference is not constant but depends on the magnitude of the stimuli. For touch the minimum perceptible difference between stimuli (Weber's fraction) is about 1/10, for hearing about 1/5, for olfaction about 1/3 and for vision about 1/100 of the magnitude of the stimuli which are being compared.

Since there is always a threshold value of the stimulus below which no sensation is produced and an upper limit beyond which pain or damage is inflicted, Weber's law is applicable only over a limited range of stimuli. Thus, although we may be able to distinguish between 20 and 21 g. by the sense of touch, we certainly cannot distinguish between 2·0 and 2·1 mg. nor between 200 and 210 kg. The relationship between sensation and stimulus may be expressed by a graph such as that shown in Fig. 36, 1, and Fechner has expressed Weber's law thus:

sensation = constant × logarithm of the stimulus.

In theory this relationship between stimulation and sensation might depend on the properties of the end-organ or on those of the central nervous system. In the case of a muscle-spindle, however, the frequency of nerve impulses set up is roughly a linear function of the logarithm of the load. A 10 per cent. increase in load produces a similar increase in the frequency of impulses in the afferent nerve, a finding that applies over a wide range of initial loads. Presumably, therefore, the sensation-stimulation relationship as described by Weber and Fechner applies to this peripheral end-organ.

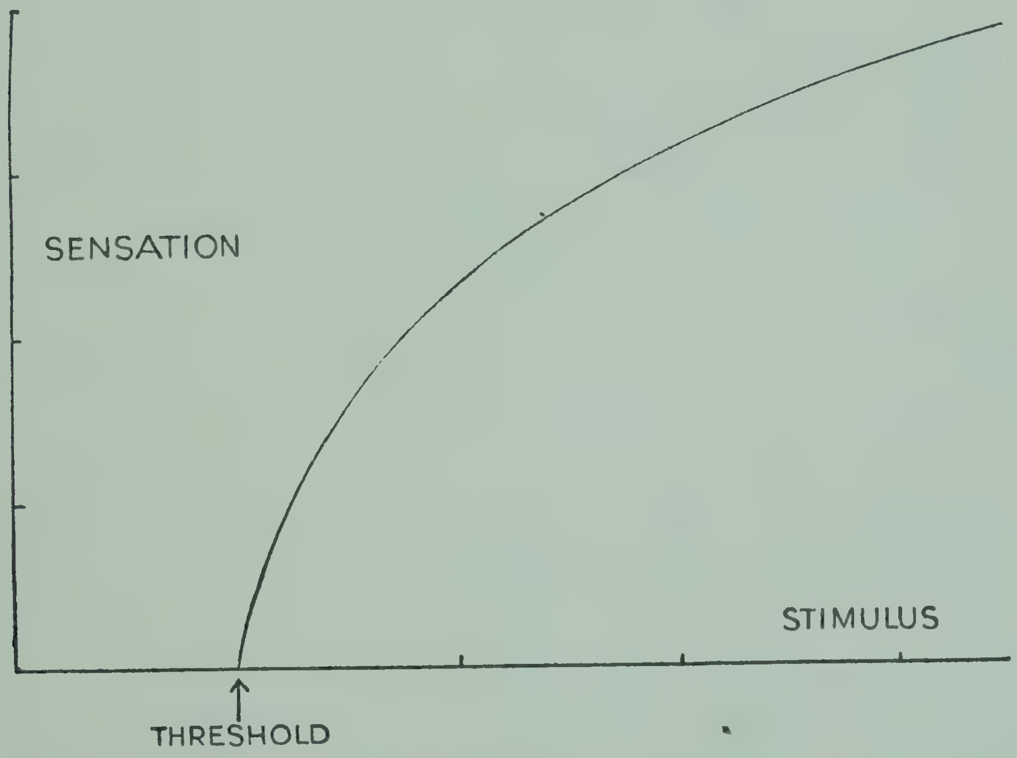


FIG. 36, 1

To illustrate the relationship between stimulus and sensation. The curve is approximately logarithmic. In an actual experiment the results would not all lie exactly on such a line but would be scattered about it.

Although the theoretical basis of the Weber-Fechner law is not clear it has been found useful especially in the physiology of hearing where the logarithmic unit, the decibel (p. 710), is in universal use. Weber's fraction for any one sense differs very much from person to person and the smallest value for any person can be realized only under special conditions.

REFERENCES

ADRIAN, E. D. (1928). *Basis of Sensation.* London: Christophers.

ADRIAN, E. D. (1932). *The Mechanism of Nervous Action.* Philadelphia: University of Pennsylvania Press.

ADRIAN, E. D. (1947). *The Physical Background of Perception.* Oxford: Clarendon Press.

GELDARD, F. A. (1953). *The Human Senses.* London: Chapman & Hall.

LEEUWEN, S. VAN (1949). The response of a frog's muscle spindle. *J. Physiol.*, **109,** 142.

MORGAN, C. T. (1943). *Physiological Psychology.* London: McGraw-Hill.

WOODWORTH, R. S., & MARQUIS, D. G. (1949). *Psychology.* London: Methuen.

CHAPTER 37

THE SKIN

THE skin separates the body from the external world and it is therefore a protective coating which prevents entry of water and microorganisms. The skin plays an important part in temperature regulation (Chapter 13), for the amount of heat lost from the body is to a large extent governed by the amount of blood flowing through the skin. There is continuously a small loss of water and salt from the body in the insensible perspiration, but such losses may become considerable when profuse sweating occurs. The excretory functions of the skin are however relatively unimportant, whereas its sensory function, the main subject of this chapter, plays an essential part in our daily lives.

STRUCTURE OF THE SKIN

Microscopic sections of the skin show two quite distinct parts, the *epidermis* on the surface and the underlying *corium* (Fig. 37, 1). The epidermis in the early embryo consists of a single layer of cells, but soon the epidermis thickens and two main parts, the *stratum corneum* and the *stratum germinativum*, can be distinguished. The stratum corneum, placed externally, has many layers of flattened cornified cells which have lost their nuclei. These dead cells are gradually cast off, but the thickness of the horny zone is maintained by new cells produced by the basal-cell layer, that is the deepest layer, of the stratum germinativum. The cells lying just superficial to the basal-cell layer are joined by fine protoplasmic bridges; when these cells are isolated the bridges break leaving little projections on each cell, hence the name prickle cells. In the thicker parts of the skin the transition from the living cells of the germinative zone to the dead cells of the stratum corneum is made prominent by two layers appropriately named *stratum granulosum* and *stratum lucidum.* In the degenerative changes which occur in this transitional zone deeply staining granules of keratohyalin appear in the cells. It seems probable that the keratohyalin granules represent a stage in the transformation of the living cell cytoplasm into keratin. When the granules have completely changed into keratin the cells assume a homogeneous appearance to form the stratum lucidum. The epidermis is renewed as fast as it is worn off and where friction and wear are greatest this layer is thickest. The pigmentation of the skin of the nipple, the mammary areola and the margins of the anus and other areas in white races is due to the presence of melanocytes or pigmentary dendritic cells lying in the stratum germinativum. These cells contain melanin granules which may also be present in the cells of the germinative layer. Hairs, nails, sweat glands and sebaceous glands are all formed by down-growths of the epidermis of the embryo.

The electron microscope shows that the dermis is made up of a network of robust collagen fibres of fairly uniform thickness showing regularly spaced cross-striations. This network is responsible for the elastic properties of the skin since comparatively few elastic fibres are to be found.

Beneath the dermis the fibrous tissue opens out and merges with the fat-containing subcutaneous tissue. The surface of the corium is formed into ridges or papillæ, projecting into the epidermis, which contain blood vessels, lymphatics and nerve endings. The blood vessels never penetrate into the epidermis but the nerve fibres reach into the germinative zone.

In certain regions of the skin, axillæ and genital regions in particular, there are *apocrine glands* each consisting of a single-coiled tube lying in the deepest part of the dermis or in the subcutaneous tissue with a duct leading to the surface which may empty into a hair follicle. These glands, which do not become active till puberty, produce an odourless viscid white milky fluid. If this is acted on by the bacteria of the skin, the decomposition products within twenty-four hours give rise to the characteristic smell of the human adult.

Sebum, the product of the sebaceous glands, can be obtained by immersing the forearm in acetone and recovering the extracted sebum. It is an amber-coloured substance resembling ear-wax made up of a mixture of triglycerides, free fatty acids, waxes, sterols, squalene and paraffins. The free fatty acids give sebum bactericidal and fungicidal properties. Irradiation of human sebum with ultra-violet light gives it antirachitic properties, probably owing to the presence of the provitamin 7-dehydrocholesterol which has been identified in the sebum of animals though not in man. If precautions to avoid contamination are taken, no fat can be extracted from the palms even when sweating is profuse. Thus, since the palms possess no sebaceous glands, it seems likely that the fat on the surface of the skin originates from the sebaceous glands alone. Sebum is produced most abundantly on the forehead, less on the trunk and very little on the extremities. Sebaceous activity is stimulated by androgens and increases at puberty. It is greater in the second half of the menstrual cycle than the first half but is very low in pregnancy.

CUTANEOUS SENSES

Stimulation of the skin gives rise to a sensation of touch, of warmth or cold, or of pain, and these are described as the four 'modalities' of cutaneous sensation. On the basis of these sensations we attribute properties to external objects.

Temperature Sense

Although we are apt to think of the skin as being uniformly sensitive all over its surface, closer examination shows that each modality has in fact a punctate distribution. In the case of the temperature sense this can be shown by passing a warm metal rod over the skin. At some places no sensation of warmth is aroused, while at other places, called 'warm spots,' a sensation of warmth is quite distinct. The 'cold spots' can be mapped out by passing a cold metal rod over the skin. When the two maps for any one area are examined it is found that the cold spots are the more numerous and that they do not coincide with the warm spots. Furthermore, tests made from day to day show that the distribution of these spots is continually altering (Fig. 37, 2). The punctate distribution is lost if the skin becomes red, as in sunburn, when all parts give rise to a sensation of warmth on being touched with a warm rod. A further observation

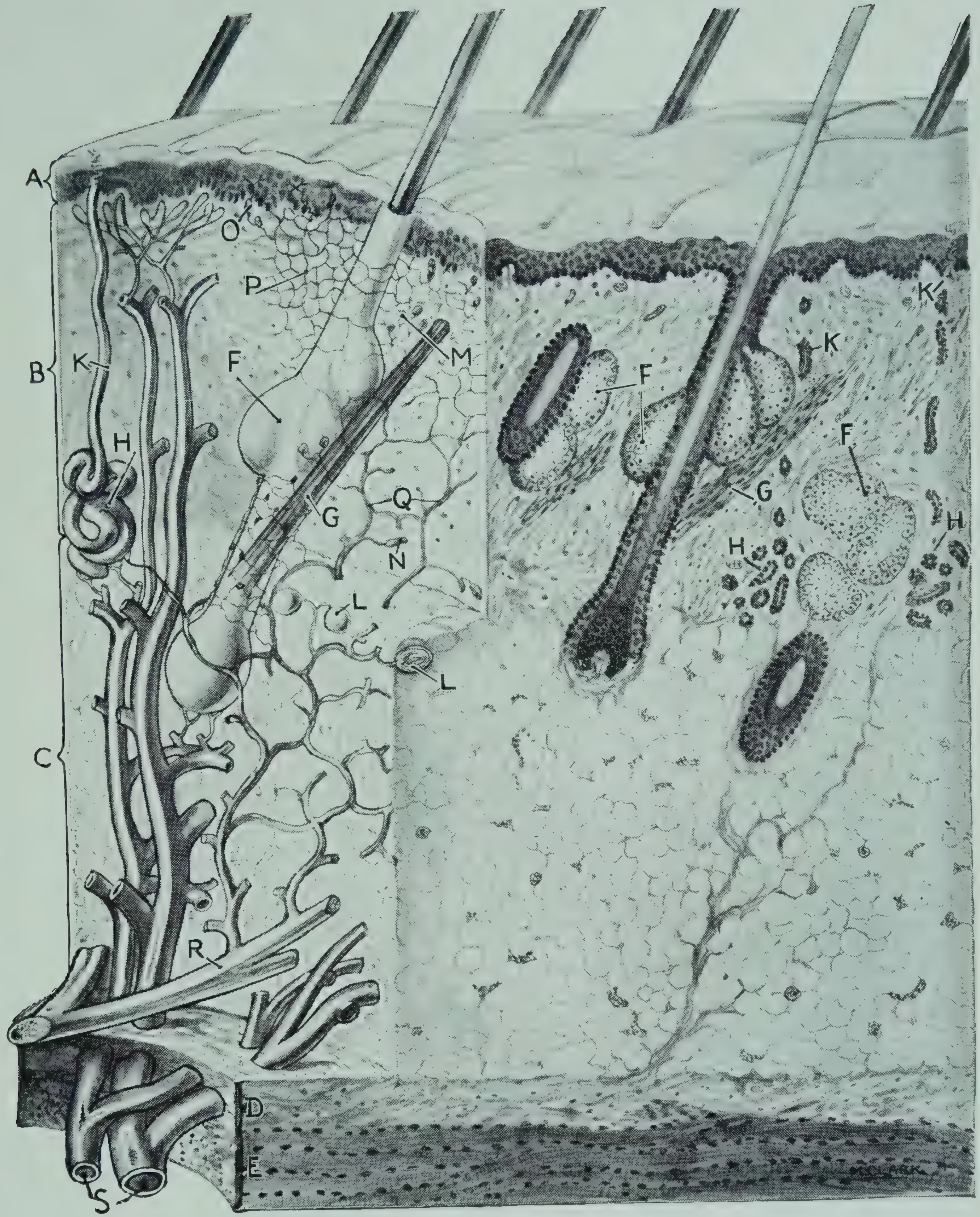


FIG. 37, 1

A scheme showing the structure of the skin.

A. Epidermis. B. Corium. C. Subcutaneous fat. D. Deep fascia. E. Muscle. F. Sebaceous glands in association with a hair follicle. G. Arrector pili muscle. H. Sweat gland. K. Duct of sweat gland. L. Pacinian corpuscle. M. Krause's end-bulb. N. Corpuscular nerve ending. O. Meissner's corpuscle in a papilla. P. Superficial nerve plexus. Q. Deep nerve plexus. R. Cutaneous nerve. S. Cutaneous vessels.

(*Gray's Anatomy*, 30th ed. p. 1249.)

suggesting that the two receptors are distinct is that a warm rod passed over a cold spot may give rise paradoxically to a sensation of cold. It has been suggested that the variability of the cold and warm spots is due to variations in the calibre of the skin vessels, and even that the vessels themselves are responsible for temperature sensations.

A piece of cloth and a piece of metal at the same temperature, either high or low, arouse quite different sensations. When touched by the finger the sensation aroused depends on the rate of withdrawal or addition of heat. A finger of one hand is placed in hot water and a finger of the other in cold; after a little both are placed in tepid water. Since this intermediate stimulus feels warm to the latter and cold to the former it is reasonable to conclude that the cause of the sensation, in this experiment, is a rise or a fall in the temperature of the nerve-endings. This cannot, however, account for the sensation of cold in the following experiment. If a cold object such as a coin is pressed on the forehead and then removed a sensation of cold persists for a little while, although at this time the temperature of the end-organ must be rising.

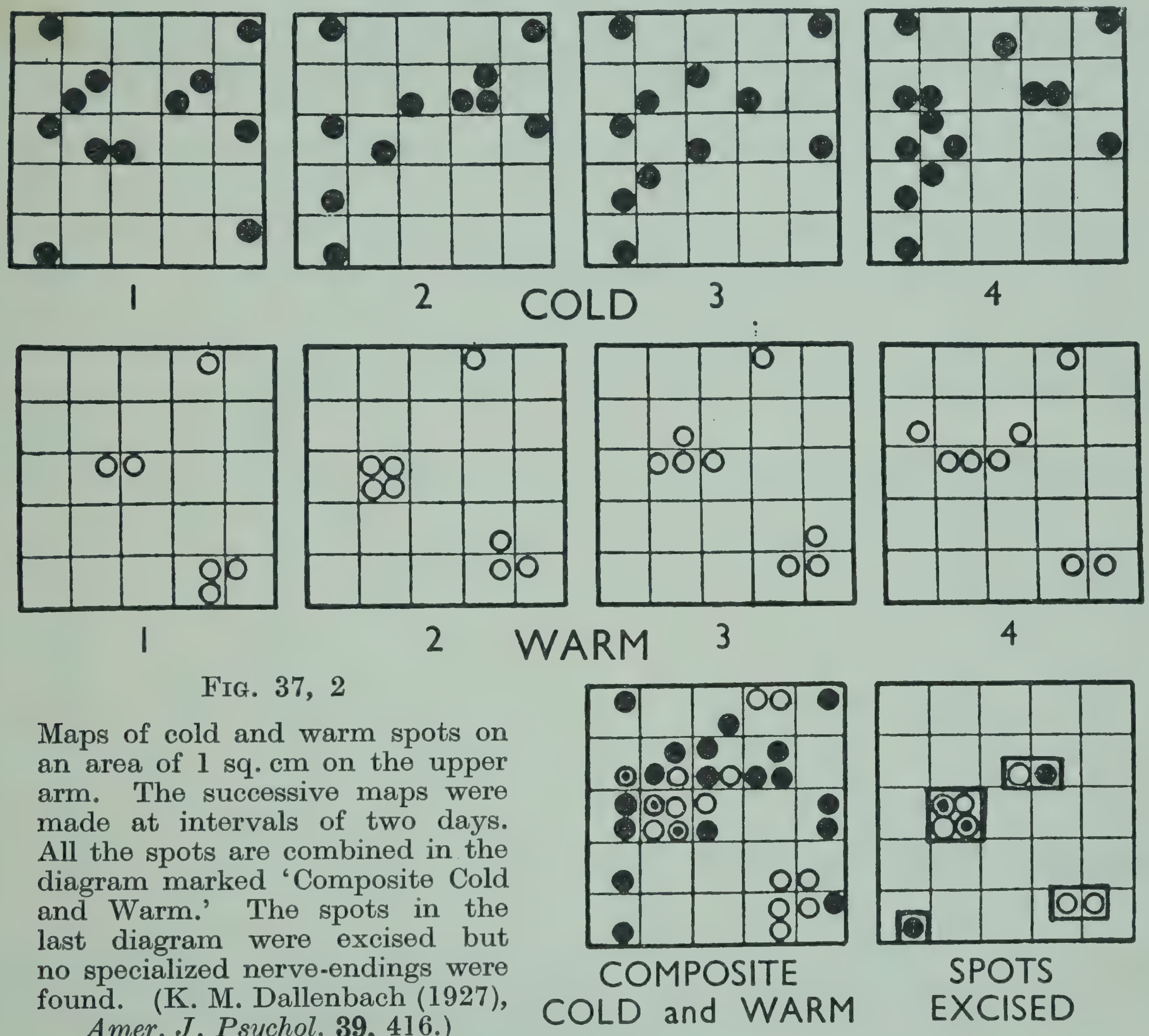


Fig. 37, 2

Maps of cold and warm spots on an area of 1 sq. cm on the upper arm. The successive maps were made at intervals of two days. All the spots are combined in the diagram marked 'Composite Cold and Warm.' The spots in the last diagram were excised but no specialized nerve-endings were found. (K. M. Dallenbach (1927), *Amer. J. Psychol.* **39**, 416.)

The experiment with the hot and cold water shows a feature common to all sensations—adaptation. The finger kept in cold water for some time does not feel cold; it becomes adjusted to the new situation which, since it arouses no sensation, becomes a new zero. The finger in the hot water also shows adaptation but in the opposite direction. This experiment also shows that the temperature sense is not to be regarded as a thermometer; with its shifting zero it cannot measure absolute temperature.

It can, however, detect quite small temperature differences. The sensitivity of the exposed areas of the skin (face, hands and scalp) is less than that of areas normally covered, such as the arms and forearms. At ordinary skin temperatures differences of more than 0·2° C can be appreciated on the arms and differences of 0·5 to 1·0° C by the fingers, the sensitivity of which is greater on the dorsal than on the palmar surface. The mucous membrane of the mouth is relatively insensitive to heat; tea can be drunk at a temperature which is painful to the finger.

THE SENSE OF TOUCH

This sense allows us to distinguish between hard and soft bodies and to judge their shape. In practice, however, muscle-joint sense may also be used to estimate hardness, and although we judge the roughness of a surface by passing the balls of the fingers gently along it yet in fact our conclusion may depend on the auditory stimuli produced at the same time. The threshold of the tactile sense is measured by von Frey's bristles which consist of a series of fibres of different thicknesses mounted on match-sticks by which they can be held between the fingers. Each is calibrated by pressing one end of the bristle on a balance till it just bends. This weight divided by the cross-sectional area of the bristle gives the pressure exerted on the skin.

TABLE 37, 3

Area	g./mm²	Lb./sq.in.
Tongue and nose	2	2·8
Lips	2·5	3·5
Finger-tip and forehead	3	4·3
Back of finger	5	7·1
Palm, arm, thigh	7	10
Forearm	8	11
Back of hand	12	17
Calf, shoulder	16	23
Abdomen	26	37
Outside of thigh	26	37
Shin and sole	28	40
Back of forearm	33	47
Loins	48	68
Thick parts of sole	250	356

Threshold pressure in g./mm² required to elicit a sensation of touch on various areas of the skin surface as measured by von Frey. In the second column the values have been converted to Lb./sq.in. to show that there is a fallacy in expressing the threshold in terms of pressure (see discussion in text).

If the skin is explored with a bristle it is found that the tactile sense is, like the other skin senses, distributed in a punctate fashion. By means of a large series of bristles von Frey found that the smallest pressure required to elicit a sensation of touch varied in different parts of the skin, as shown in g./mm² in the middle column of Table 37, 3. If these values are converted into Lb./sq.in. (third column of Table 37, 3) quite ridiculous values are obtained; the lips (second entry in the table) are certainly sensitive to much less than 3·5 Lb. on an area of 1 sq.in. Thus the adequate stimulus for eliciting a sensation of touch cannot be mere pressure, but probably involves skin deformation, as is shown by placing a finger in a bowl of mercury. Contact is felt only at the air/mercury junction where the skin is deformed by the abrupt change of pressure, and not deeper in the mercury at the tip of the finger where, although the pressure is actually much greater, there is no deformation.

The touch spots are thickly distributed around hair follicles and the root of each hair is surrounded by a network of nerve fibres. Each hair pivoted at the skin surface acts as a lever to transmit pressure to the nerve-endings round the shaft near the root. When the hairs are removed the sensitivity to touch is greatly reduced.

It is usual to talk of touch ‘spots’ as if they were separate and discrete areas but it is found both by histological and by electrophysiological methods that each single nerve fibre innervates several square millimetres, and that each such area contains endings of other fibres. A single contact with a bristle may therefore send nerve impulses into the spinal cord by more than one route.

Stability of Sensory Spots.—The map of the distribution of the touch spots in any given area varies from day to day, but if the observations are made with a careful technique on trained subjects the pattern of the touch spots becomes more nearly constant than in a poorly managed experiment. This variability in position of the touch spots makes it difficult to believe in an anatomical, that is a fixed, basis for the touch sense. The sensitivity of the skin varies from place to place on any one occasion, and a strong bristle reveals more touch spots than a weak one. If, in addition, the sensitivity of any one place varied from day to day it would be easy to understand both the variability of the position of the touch spots and the apparently punctate distribution in the absence of discrete fixed end-organs.

The tactile sensation is unique in lasting only a very little longer than the stimulus. If a finger is held lightly against a revolving cog wheel each stimulus gives rise to a separate sensation until the contacts follow one another at more than 500 or 600 per second. Above this frequency of stimulation the rotating cog wheel is described as smooth, the separate sensations having fused into a continuous sensation. If, however, the amplitude of vibration of an object applied to the skin is very high, a sensation of vibration may persist up to 8000 c/s and, with practice, amplified speech vibrations can be felt by the fingers and recognized. The speech vibrations may, however, be recognized by their intensity rather than by their frequency, and the ability to do so may be of some value to a deaf person.

The tactile sense shows adaptation which is partly peripheral and partly central. We are aware of the contact of our clothes with the skin when they are first put on but this sensation quickly disappears. It can, however, be brought to consciousness by an act of the will or by a slight movement of the body. In the same way a new denture is very obvious at first but after a time the wearer becomes unconscious of its presence.

Localization.—When a blindfolded person is touched, not only does he have an impression of the strength of the stimulus but he can also indicate the point touched with a considerable degree of accuracy. The power of localization depends partly on the position at which the nerve fibres from the tactile end-organs enter the spinal cord and on their higher connexions, and partly on experience. Aristotle's experiment shows the influence of experience. A pencil laid between two crossed fingers of a blindfolded observer gives the sensation of contact by two pencils. Experience also plays a part in visual localization. After removal of a congenital cataract in persons blind from birth the patient is not able at first to localize his visual impressions; the power to do so grows as experience of his new sense develops.

Discrimination.—If two points on the skin are touched simultaneously and with approximately the same degree of pressure by the points of a pair of dividers (æsthesiometer) the subject will report that two separate points have been touched, provided that the distance exceeds a value depending on the area of skin under test (Table 37, 4). If the skin is carefully explored for touch spots it is found that in the area showing the lowest two-point threshold they are less than 1 mm apart but at no part of the body is the distance between these spots more than 6 mm. Scarcity of touch spots is not the explanation of the poor discrimination of the less mobile parts of the body surface.

TABLE 37, 4

The Two-Point Threshold *

Area	Separation in mm
Tip of tongue	1
Anterior surface of finger-tip	2
Posterior surface of third phalanx	6
Palm of hand	11
Back of hand	32
Back of neck	54
Middle of back	67

* The distance at which two points on the skin touched simultaneously with approximately the same pressure must lie apart to be recognized as separate points of contact.

This two-point threshold is usually three or four times the error of localization in the same region. This apparent anomaly is probably due to the diffuse nature of the stimulus. A blunt point on the skin stretches the surrounding skin and, according to Woodworth, localization is probably determined by the centre of strongest stimulation; in the case of two-point stimulation the whole stretched area is effective in giving a broad impression and obscuring the presence of the two points.

Projection.—This is most highly developed in vision but it occurs also with the tactile sensation. If a pudding is stirred with a spoon, lumps may be detected and the bottom of the pan feels hard. The stimuli which originate these sensations are applied to the skin in contact with the upper end of the spoon but the sensations are projected to the lower end of the spoon.

THE PAIN SENSE

Pain is often referred or projected to the surface of the body when internal organs are diseased (p. 805) but here we are concerned only with pain actually arising in the skin. Painful stimuli or, as Sherrington called them, nocuous stimuli call forth withdrawal reflexes which occur even when the brain is removed from an animal and only the spinal cord remains intact. In the intact animal the same reflexes occur and it is reasonable to suppose that a painful sensation is evoked as in man.

Pain elicited by stimulation of the skin has a pricking, burning or itching quality and is well localized. In sunburnt skin the pain threshold may be lowered by half; or it may be raised by one third by distracting the subject's attention. If the finger just below the nail-bed is pressed against a hot electric light bulb a double pain is produced. The first sensation arises abruptly and is pricking in quality and is probably carried in relatively large fibres at about 10 metres per second. The second sensation follows a second or more later, has a burning quality but is less abrupt in onset and disappearance; it is probably carried in non-myelinated fibres. Since asphyxia of the skin abolishes the first pain but not the second, the underlying mechanisms must be different. Various observers have produced evidence to show that pain is carried in A, B and C fibres (see p. 743).

TABLE 37, 5

Area	Threshold g./mm^2
Cornea	0·2
Conjunctiva	2
Abdomen	15
Front of forearm	20
Back of forearm	30
Calf of leg	30
Back of hand	100
Sole	200
Finger-tip	300

The minimum stimulus required to elicit a sensation of pain at various parts of the body. Note the want of parallelism with the figures for the sense of touch in Table 37, 3 (von Frey).

It could be asserted that excessive stimulation of any sense organ is apt to cause pain but quite gentle exploration of the skin surface with a needle produces a painful sensation at certain places. Once again the pain end-organs seem to be distributed in a punctate fashion, independently of the end-organs of touch or temperature. Any given area of skin may show ten times as many pain spots as touch spots. These pain endings have different characteristics from the tactile end-organs; the threshold stimulus (Table 37, 5) is quite different and the order of increasing sensitivity is completely different from that given for touch in Table 37, 3. The pain sense has a relatively slow rate of adaptation. Asphyxia reduces the response of the other senses but increases the response to painful stimuli. In the disease of the spinal cord called syringomyelia pain may disappear while the other cutaneous senses remain. There is no doubt therefore that pain is a separate modality and that it is not simply due to excessive stimulation of nerve-endings.

It is difficult to say how a stimulus gives rise to pain. Presumably it depolarizes the end-organ or suddenly increases its permeability. Mechanical or chemical stimuli could no doubt produce either effect directly or perhaps indirectly by releasing from damaged cells a chemical substance. Keele and his colleagues are investigating this possibility by the simple expedient of raising a skin blister with cantharidin and then applying the various substances to be tested to the exposed base of the blister. Blood, plasma or inflammatory exudate withdrawn into a siliconed syringe have very little pain-producing activity; on transfer to a glass vessel, pain-producing activity appears rapidly and then declines. The active agent behaves like a polypeptide. It is reasonable to speculate that tissue injury might, by allowing plasma to escape from the capillaries, release a polypeptide and cause pain but damage to blood platelets could also release 5-hydroxytryptamine (p. 946), another very potent pain-producing substance. Histamine applied to the blister base causes pain but only in high concentrations unlikely to be produced by injury.

Pruritus or itching may be defined as 'an unpleasant cutaneous sensation which provokes the desire to scratch.' The sensation may be mild, in which case it is almost pleasurable, and this is quite common in elderly people; sometimes it may be so severe as to cause the sufferer to attempt suicide. Pruritus may be generalized, suggesting a metabolic cause, or confined to quite a small area of skin or to a particular part of the body such as, commonly, the perineum or the nostrils. The sensation arises in nerve-endings in the epidermis since pruritus cannot be elicited in areas denuded of epidermis. It can be produced in a sensitive area by chemical, mechanical, thermal or electrical stimuli, and mild forms of such stimuli may elicit itching in a localized area of skin long after the initial pruritus has subsided. Thus, light stroking of an insect bite may produce itching weeks after the local reaction has subsided. In other words, the skin remains in a state of increased excitability. If the itching so produced leads to vigorous scratching, the area of excitable skin enlarges and the threshold is further reduced.

Pruritus can be produced experimentally by the introduction of histamine into the skin, and in this case the sensation is much diminished or abolished by prior administration of anti-histamine substances, for example mepyramine maleate. This suggests that the pruritus like the triple response is produced by a histamine-like substance released in the skin. In disease, when it is not infrequently accompanied by redness and wealing of the skin, the sensation may be abolished by prior administration of anti-histamine substances; but this is not always so, for example the itching in jaundice is not abolished by mepyramine. It is unlikely, therefore, that pruritus is always produced by histamine. There is no doubt that pruritus is often initiated or aggravated by psychic disturbances.

Lewis in 1936 claimed, on the basis of experiments on man, that there was another system of nerve fibres in the skin, which he termed the nocifensor system. Within a few seconds after making a small crush on the skin—the forearm is a suitable place—a small area round the crush shows tenderness (hyperalgesia), and in some persons this hyperalgesia spreads gradually over a considerable area which remains tender for an hour or so. If the skin is anæsthetized before the crush is made, tenderness does not develop until sensation returns. This suggests that nerves lying in the skin, so-called nocifensor nerves, must be responsible for the phenomenon, and that it is not the pain of the crush which causes the spreading tenderness, but the products of injury. Such a separate system of nerves has not been demonstrated histologically.

Anatomical Basis of Cutaneous Sensations

The experiments which have just been described seem to provide sufficient evidence for the recognition of four 'modalities' of cutaneous sensation, and the punctate distribution of the sensitive areas makes it logical to search for special receptors. Indeed the older histologists found many organized (that is elaborate) endings in the skin and, as it appears now, rather rashly labelled them 'pressure' endings, 'cold' endings and so on, simply on the basis of their morphology. Since then many workers, having found a spot on the skin consistently sensitive to a particular sensation, marked it, excised it and examined it histologically. They have been almost invariably disappointed and have found no organized endings in spite of the use of a great variety of histological techniques. A group of workers at Oxford have recently compared several areas of skin both experimentally and histologically. It was found that there was no important difference in sensitivity to each of the four modalities of sensation of the skin of the auricle (external ear), of the front of the forearm and the palmar aspect of the fingers. A complicated nerve plexus was found in excised skin of the auricle ending either in a basket-like network around hair follicles or by repeated breaking-up into beaded branches which ended freely in the superficial layer of the dermis and deeper layers of the epidermis. No organized endings were seen, although the same techniques applied to the skin of the fingers revealed many organized endings as well as free terminations. Thus in the auricle at least, organized endings are not necessary for the perception of touch, pain, cold and warm stimuli, and this applies to all hairy skin which has been investigated. This does not mean that the organized endings, where they occur, are functionless. The older anatomists described distinct varieties of organized endings, but closer inspection of a large number has shown many intermediate forms which do not permit easy morphological classification. In both hairless and hairy regions there are free nerve terminals; as we pass from a hairless area (the lip mucosa for example) to a hairy area (the facial skin) the organized endings disappear and give way to basket-like endings around hair follicles. Until some revolutionary histological technique is discovered it would be wise to drop the search for an anatomical basis for the four modalities. And finally it must be admitted that excessive preoccupation with the outmoded notion that function is always necessarily dependent upon specific structure leads to sterile controversy of which the above is a good example.

REFERENCES

Armstrong, D., Keele, C. A., Jepson, J. B. & Stewart, J. W. (1954). Development of pain-producing substance in human plasma. *Nature*, **174**, 791–792. London.

Carruthers, C. & Suntzeff, V. (1954). Biochemistry and physiology of epidermis. *Physiol. Rev.* **34**, 229–243.

Chase, H. B. (1954). Growth of the hair. *Physiol. Rev.* **34**, 113–126.

Hagen, E., Knoche, H., Sinclair, D. C. & Weddell, G. (1953). The role of specialized nerve terminals in cutaneous sensibility. *Proc. roy. Soc.* B, **141**, 279–287.

Hamilton, J. B. (Ed.) (1951). The growth, replacement and types of hair. *Ann. N.Y. Acad. Sci.* **53**, 461–752.

Hardy, J. D., Wolff, H. G. & Goodell, H. (1952). *Pain Sensations and Reactions.* Baltimore: Williams & Wilkins.

Mackenna, R. M. B. (1953–4). The scientific approach to dermatology. *Lect. Sci. Bas. Med.* **3**, 347–371.

Rothman, S. (1954). *Physiology and Biochemistry of the Skin.* Chicago University Press.

Tunbridge, R. E., Tattersall, R. N., Hall, D. A., Astbury, W. T. & Reed, R. (1952). The fibrous structure of normal and abnormal human skin. *Clin. Sci.* **11**, 315–331.

Wheatley, V. R. (1953). Some aspects of the nature and function of sebum. *St. Bart. Hosp. J.* **57**, 5–9.

Woodworth, R. S. (1950). *Experimental Psychology.* London: Methuen.

CHAPTER 38

THE CHEMICAL SENSES

TASTE and smell are usually referred to as the chemical senses because the receptors are stimulated normally only by contact with chemical substances. In addition the nose, mouth and genital apertures, and many mucous membranes, are sensitive to irritants such as ammonia. This third form of chemical sense is referred to as 'common chemical sensibility.' A sensation described as a taste may in fact arise from the simultaneous stimulation of the taste and smell receptors; this is the case with onions and with wine. The close association between the two senses is shown by the lack of flavour of many food substances when the sense of smell is lost during a cold in the head. It is convenient to use the word flavour for the sensation aroused by the simultaneous stimulation of these receptors.

The chemical senses of the higher animals may be regarded as the representation in the evolutionary process of that property of the protozoa which brings about positive or negative chemotaxis, that is, the movement towards food or away from harmful substances. The senses of taste and smell are not well developed in man and indeed they might be described as vestigial. Man has come to depend more on his visual equipment and it is, therefore, not surprising to find that the olfactory part of his brain is relatively much smaller than that of many of the lower animals. The sense of smell is normally projected on to the environment but this does not occur with the sense of taste. Although an odorous substance stimulates end-organs in the nose yet the sensation is referred to or projected on to the source of the odour. Thus we say that the rose has a sweet perfume, not 'there is a sweet perfume in the nose.' Similarly in vision we refer the sensation to the external world and not to the retina which is the end-organ.

SENSE OF SMELL (OLFACTION)

The greater part (or respiratory region) of the nasal cavities is lined by columnar ciliated epithelium interspersed with goblet cells which are capable of producing large quantities of mucus. In man the olfactory area, about half a square inch in area on either side, is situated in a narrow cleft in the highest part of each nasal cavity above the superior concha. This area possesses olfactory cells and supporting cells (Fig. 38, 1). The latter cells, which contain granules of yellow pigment, are columnar near the surface and taper off into fine processes which pass into the deeper parts of the mucous membrane. The olfactory cells are bipolar and lie among the tapered ends of the supporting cells; each has one or more thin olfactory hairs passing to the surface of the olfactory area and a centrally directed process or olfactory nerve fibre. These non-medullated nerve fibres form a plexus in the deeper parts of the mucosa and then pass through the cribriform plate of the ethmoid bone to the olfactory bulb.

Tubular glands (nasal glands) lying below the olfactory epithelium keep it moist by secreting a clear mucous fluid which passes by ducts to the surface.

There is an association between pigmentation and the sense of smell. In animals in which the sense of smell is well developed the olfactory area is often brown and the secretion of the nasal glands may be brown also; on the other hand, albino animals are liable to be poisoned because they cannot identify poisonous plants by their odour. The dark-skinned races are said to possess a much more acute sense of smell than white people.

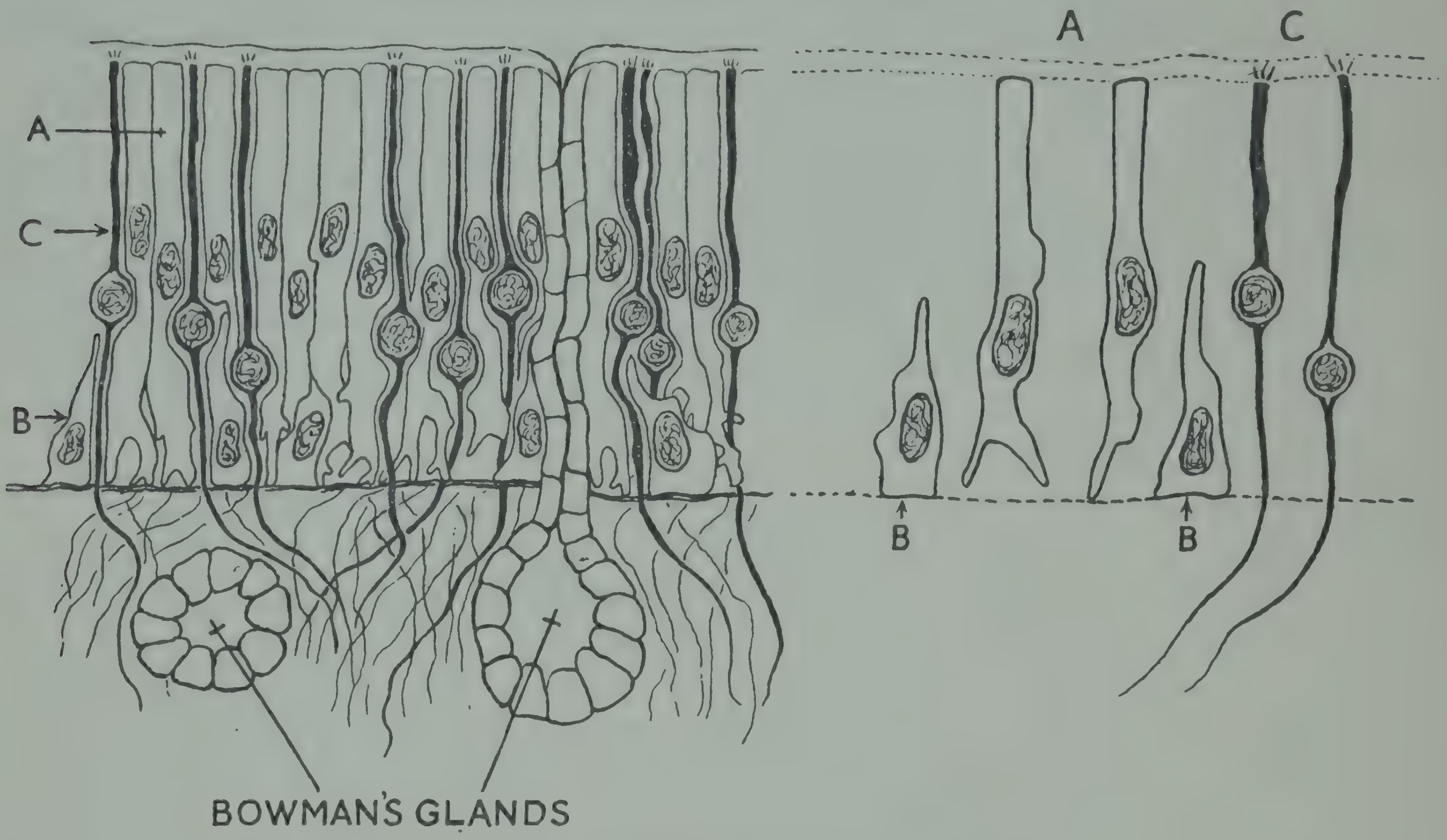


FIG. 38, 1

Olfactory epithelium and component cells. A, sustentacular; B, basal; C, olfactory. The nasal glands (Bowman's glands) keep the epithelium moist. There are several million olfactory cells (C) in the olfactory area in man. (*After* H. S. D. Garven.)

The route taken by the stream of air through the nose has been demonstrated in several ways. For example the nasal mucosa of a cadaver was covered with red litmus and air containing ammonia was blown through the nose in imitation of inspiration or expiration. The air pathway was shown by a blue colour. Again, direct inspection of the nasal cavities in the living subject can be carried out after inhalation of light magnesia powder. These methods show that the air stream does not rise above the superior concha on inspiration and is at an even lower level on expiration. Thus in ordinary quiet breathing the air does not impinge on the olfactory area directly but only as eddies coming off the main stream. When we attempt to smell a substance by sniffing, the air-stream reaches the olfactory area directly either because it is travelling more rapidly or because it is directed there by a change in the shape of the nose.

It is by no means certain how the olfactory cells are stimulated. It is clear, however, that if air cannot reach the olfactory areas, for example during a cold in the head, then the sense of smell is lost. If olfaction is the modern representation of the primitive chemical sense of single-celled organisms then it might be supposed that the odorous substances go into solution in the moist layer covering the olfactory area. However, a saline solution of an odoriferous substance placed in the nose may give rise to quite a different sensation from that usually regarded as the characteristic odour of the substance. Another difficulty is that some odoriferous substances are not water-soluble. Many of these are, however, soluble in lipids and it may be that this property allows them to penetrate into the olfactory cells and excite them.

Many substances appear to lose their odour within a very few minutes of presentation. This may be a result of either a rapid adaptation of the end-organs or a washing away of the odoriferous particles by mucus. Adaptation is at least part of the explanation since loss of odour tends to be specific, that is, any other odour, if quite different from the first, is easily perceived. The inability to perceive odours after prolonged exposure should perhaps be termed fatigue rather than adaptation. The former term seems more appropriate in the case where the occupants of a stuffy room are quite unaware of their odour whereas a newcomer is at once unpleasantly conscious of it. Neutralization of one odour by another occurs to some extent, but in practice disagreeable odours, for example that of fæces, are disguised by overwhelming them with powerful, but more pleasant, odorous substances.

Possibly because of the vestigial character of the olfactory sense in man no satisfactory classification of odours has been drawn up. One of the simplest suggestions is that there are four main types, (1) fragrant, (2) acid or sour, (3) burnt and (4) caprylic (disagreeable or putrid), but this classification is much less satisfactory than that for taste (p. 665). There is no simple relationship between chemical constitution and odour. Similar odours may be given by substances with very different chemical constitution, for example hydrocyanic acid and nitrobenzene both smell of bitter almonds. In the case of optically active substances it frequently happens that one stereoisomer has a more intense smell than its mirror image.

The minimum stimulus required to arouse a sensation of smell varies with different substances, but there are some substances, like mercaptan, chlorophenol and skatole, which can be detected in the air at very high dilutions. It is possible to detect 0·01 mg. of mercaptan in a room of 230 cubic metres (say 28 ft. by 22 ft. by 13 ft. high). Since 0·01 mg. of mercaptan contains 1×10^{17} molecules each sniff of 20 ml. contains 1×10^{10} molecules. If all these are absorbed on the surface layer of the olfactory epithelium then there will be quite a high concentration, but in the absence of some selective action by the olfactory epithelium it is difficult to understand how one molecule of mercaptan in fifty thousand molecules of air can produce an olfactory sensation.

Adrian has made records of the electrical activity in the olfactory tract of the rabbit and has found that, if the spontaneous discharge of the cells in the olfactory bulb has been suppressed by a general anæsthetic, discharges in the olfactory tract can be detected when air containing an odorous substance is drawn through the nose. The anterior part of the olfactory area is easily excited by esters (fruity smell) which are slightly soluble in water. Oil-soluble substances on the other hand excite the posterior area more easily. The discharge in the latter case has a longer latent period, begins more slowly and dies away more slowly, than in the former.

Moncrieff has recently found by making tests of adsorption of odoriferous substances on a series of adsorbents, including active carbon and alumina and fuller's earth, that substances having similar odours are adsorbed to similar extent on the adsorbents even if they have quite different chemical constitutions; conversely, substances having different odours behave differently towards the adsorbents used. It is a considerable step from these experiments to a theory of olfaction, but it is tempting to suppose that some sort of selective adsorption goes on in the nose.

SENSE OF TASTE (GUSTATION)

The taste receptors or *taste buds*, about 9000 in the adult, are placed mainly on the peripheral parts of the dorsum of the tongue, being most easily found in the groove surrounding the vallate papillæ: some buds occur on the soft palate and a few on the epiglottis. In the child they are much more numerous and much more widely distributed over the tongue and on the insides of the cheeks. The oval taste buds embedded in the stratified epithelium covering the tongue contain thin fusiform taste cells surrounded by supporting cells like the staves of a barrel. Each taste cell ends in a gustatory hair which passes up towards the minute opening of the taste bud (gustatory pore) upon the surface (Fig. 38, 2).

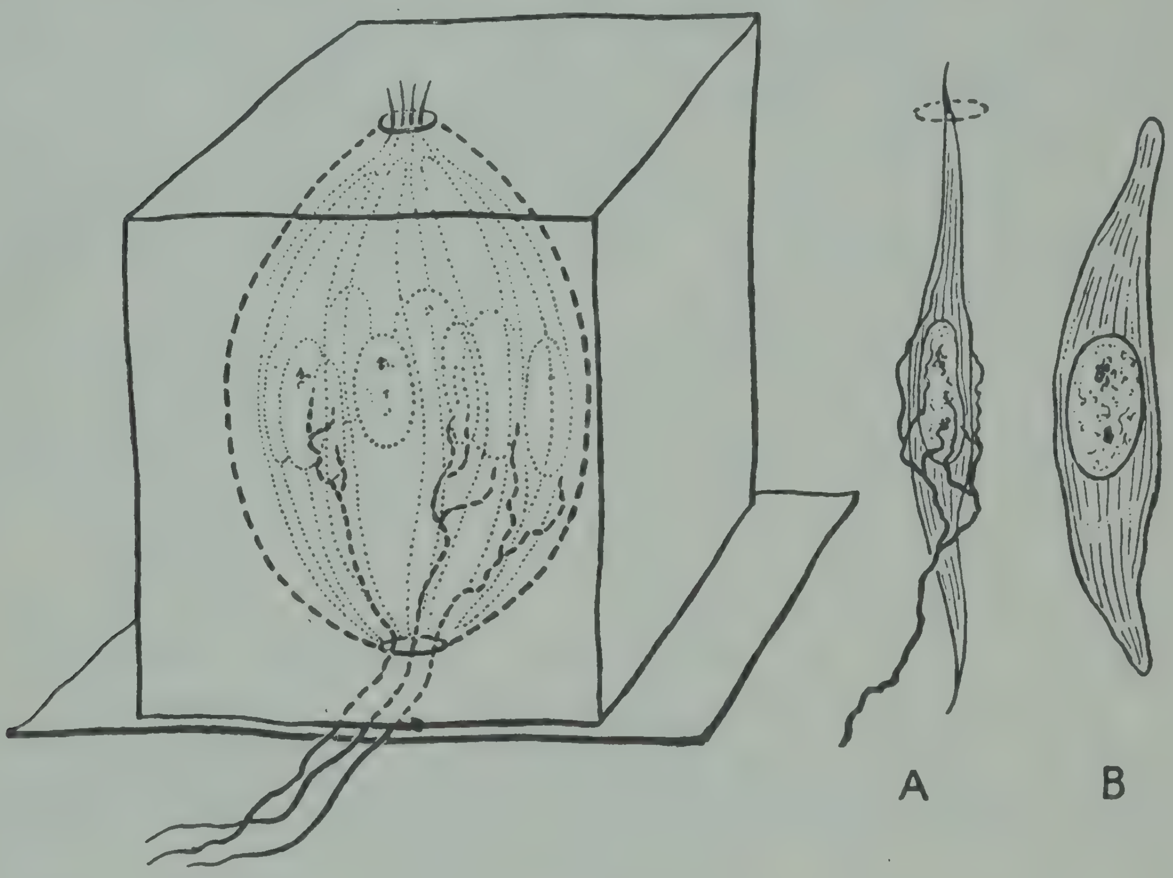


FIG. 38, 2

A small cube of epithelium from the sulcus surrounding a vallate papilla, showing a taste bud with external and internal pores and nerve fibres. A, gustatory cell with nerve fibres. B, sustentacular cell. (*After* H. S. D. Garven.)

The *nerves of taste* are the lingual branch of the glossopharyngeal (IX) for the posterior third of the tongue, including the vallate papillæ, while the anterior two-thirds are served by the chorda tympani branch of the facial (VII). The greater superficial petrosal nerve supplies gustatory fibres to the palate. Taste buds near the epiglottis are innervated by the vagus. All the nerve fibres pass centrally to the tractus solitarius, a long column of grey matter in the medulla oblongata (Fig. 38, 3). From here fibres cross the mid-line and pass to the posterior ventral nucleus of the thalamus and then on to the lower part of the postcentral gyrus and probably other areas nearby. The lingual branch of the Vth nerve is the nerve of common sensibility for the anterior two-thirds of the tongue; the lingual branch of the glossopharyngeal carries both taste and common sensibility fibres for the posterior third.

Probably no true taste fibres pass into the brain stem by the fifth nerve but tactile, pain and thermal sensations conveyed by this nerve play an important part in the recognition of taste sensations. The integration of taste sensations of the chorda with common sensibility conveyed by the fifth nerve provides a full gustatory experience. When the fifth nerve in man is cut or destroyed the sense of taste is immediately lost in the front of the tongue but it returns in some cases after a few hours, in others after years. There is no such separation of routes for taste and common sensibility from the posterior part of the tongue as both enter the brain stem by the glossopharyngeal nerve. If the facial nerve is divided above its ganglion the sensation of taste is lost permanently on both the palate and the anterior part of the tongue; although the fifth nerve is intact, it is unable to provide a sensation of taste.

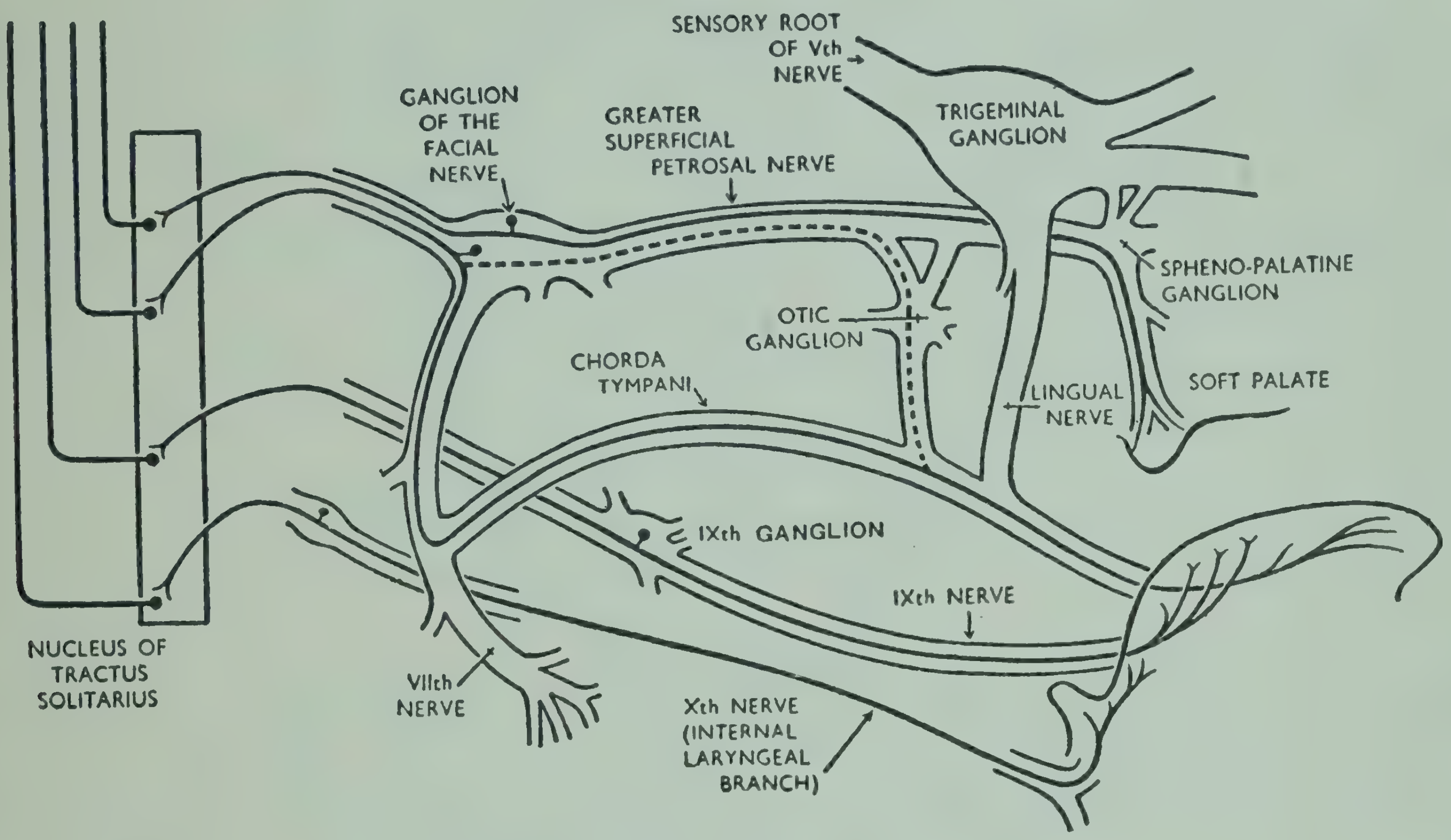

SENSORY ROOT OF Vth NERVE
GANGLION OF THE FACIAL NERVE
GREATER SUPERFICIAL PETROSAL NERVE
TRIGEMINAL GANGLION
SPHENO-PALATINE GANGLION
OTIC GANGLION
CHORDA TYMPANI
LINGUAL NERVE
SOFT PALATE
IXth GANGLION
IXth NERVE
NUCLEUS OF TRACTUS SOLITARIUS
VIIth NERVE
Xth NERVE (INTERNAL LARYNGEAL BRANCH)

Fig. 38, 3

Pathways of the taste fibres are given by bold lines. The interrupted line is an alternative route for taste from the anterior two-thirds of the tongue which appears to be present in some persons. Although these pathways are agreed to by most neurologists there is still some dubiety as to details. Diagram modified and simplified from A. Brodal (1948), *Neurological Anatomy in relation to Clinical Medicine.* Oxford: Clarendon Press.

The sensibility of the tongue is investigated by placing small drops of the test substances in solution in water on the protruded tongue while the subject holds his nose. Such experiments show that there are only four tastes, namely *sweet, sour, salt* and *bitter*. Sweet tastes are most easily perceived at the tip of the tongue, bitter at the back, sour at the edge and salt both on the tip and the edge (Fig. 38, 4). It is not possible to test single taste buds as they are arranged in groups. These groups are found mainly on the walls of the vallate papillæ and to a lesser extent upon those of the fungiform papillæ. Some papillæ respond only to sweet solutions and others only to sour, but most appear to respond to all kinds of gustatory stimuli. In such experiments the speed of recognition of the particular variety of taste, especially when the tongue is protruded, is relatively slow. The full sensation allowing the taste to be correctly recognized may not appear for several seconds. This is in complete contrast to light and auditory stimuli which are recognized after a very short latent period. Tastes in everyday life, that is apart from laboratory experiments, are generally a mixture of gustatory, tactile, thermal, kinæsthetic and olfactory sensations. By kinæsthetic sensation is meant the notion of hardness or toughness or elasticity obtained from the impulses arising in the end-organs of the muscles and joints used in chewing. If the nose is held and chewing is forbidden it is exceedingly difficult to distinguish between turnips, apples and onions.

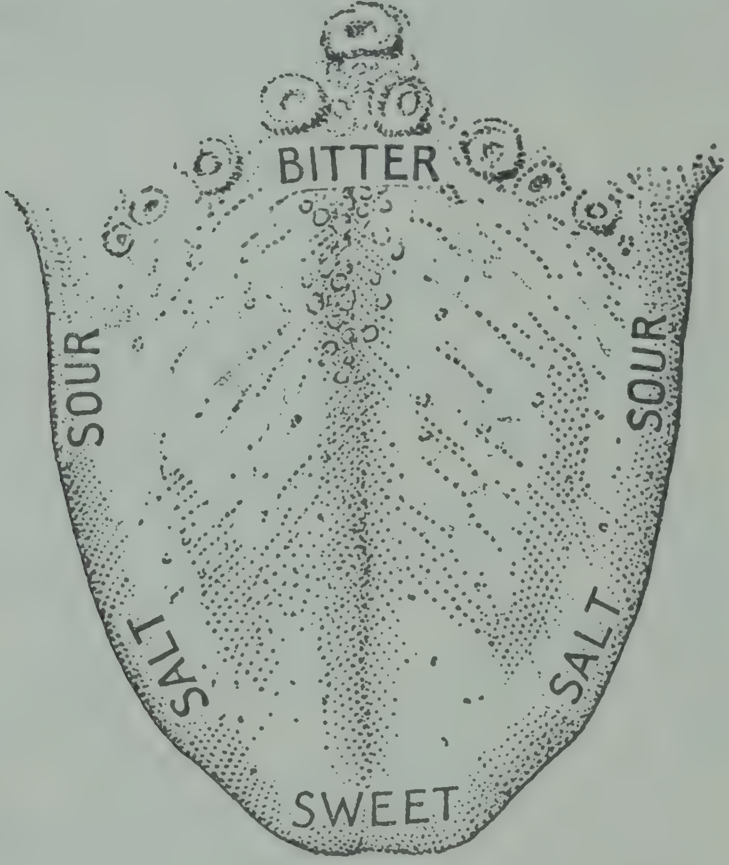


Fig. 38, 4

The areas of the tongue most sensitive to the various gustatory stimuli are indicated. The centre of the dorsum is not sensitive to such stimuli but only to tactile, thermal and painful stimulation.

Sour-tasting materials all contain acids—vinegar contains acetic acid, citrus fruit citric acid, sour milk lactic acid. Whereas very weak acids do not have a sour taste, strong acids like hydrochloric acid, which dissociates almost completely in aqueous solution, have a sour taste at very high dilution. Obviously the sour taste of hydrochloric acid might be due to either H^+ or Cl^- ions. Now sodium chloride dissociates almost completely into Na^+ and Cl^- ions, but 0·00125 N-NaCl is tasteless whereas 0·00125 N-HCl is sour. In this case the sourness must be due to the hydrogen ions. Organic acids like citric acid are, however, more sour than would be expected if hydrogen-ion concentration were the only factor determining sourness. It seems reasonable there

and the undissociated molecule as well as the

determining the degree of sourness of an acid.

The typical salt substance is sodium chloride; calcium chloride potassium chloride also taste salt but since all chlorides do not taste exactly the same it seems that the cations must have a modifying influence. The salt taste is not, however, a specific property of the chloride ion since sodium bromide, sodium iodide, sodium sulphate and sodium nitrate all have salt tastes.

The factors determining sweetness are more puzzling still, and are not clearly related to chemical constitution. A number of organic substances, such as sucrose, glycerine and saccharine, and a few inorganic substances, such as the salts of lead and beryllium, all have a sweet taste. The α-amino-acids are sweet but polypeptides are often bitter. The sweetest sugar is fructose, then come sucrose and glucose, with lactose a long way behind. If a 2 per cent. solution of sucrose is used as the standard of comparison then saccharine is apparently 700 times sweeter than sucrose.

A bitter taste is given by a variety of substances, for example cæsium chloride, potassium iodide, magnesium sulphate, picric acid, and alkaloids such as quinine, strychnine and aloin.

Some tastes are not easily included in the above fourfold classification, for example alkaline and metallic tastes. It may be that in these cases the common chemical sensibility is stimulated. Freshly distilled water has an alkaline taste owing to the absence of carbon dioxide. Some substances give rise to a persistent taste sensation, for example, saccharine and iodides. This is presumably due to the fact that they are excreted in the saliva.

Thus with the exception of the sour taste attempts to correlate chemical constitution with gustatory sensation have failed just as completely as in olfaction. To illustrate just how baffling the problem is, it may be mentioned that L-tryptophan is bitter, DL-tryptophan is sweet and D-tryptophan is nearly tasteless.

Evidence for the separate identity of the four tastes is difficult to obtain. No instance has been reported in which one of the four taste sensations has been absent or impaired by disease. It has, however, been found that phenylthiocarbamide and allied substances are bitter to about 60 per cent. of persons and tasteless (unless very high concentrations are used) to 40 per cent., and that the ability to taste these substances is inherited according to Mendelian laws. Most people cannot taste sodium benzoate, but a few describe it either as sweet or bitter.

The threshold concentration needed to arouse a sensation of taste varies widely with different individuals and depends somewhat on the method of testing but the following are average values. Sour, 0·0045 per cent. HCl; salt, 0·055 per cent. NaCl; sweet, 0·45 per cent. fructose, 0·7 per cent. sucrose, 0·001 per cent. saccharine; bitter, 0·001 per cent. brucine. The smallest difference in the strength of a test solution which can be distinguished is about 30 per cent.

It has been shown by histological methods that the epithelium of the folds between the papillæ contains numerous enzymes including alkaline glycerophosphatase and esterases. Vanillin, infusion of tea and other substances known to have a taste inhibit the phosphatase activity; quinine ut affecting the phosphatase activity; have no effect on these enzymes. Baradi and tentatively that a substance which has a characteristic stimulates or depresses enzyme activity in a definite pattern which is recognized in some way by the central nervous system. If this is true the status of the taste buds as the sole organs of taste may have to be reconsidered.

The sense of taste shows adaptation. After cane sugar has been held in the mouth for two minutes the threshold for cane sugar may be raised ten times. At the same time the threshold for salt may be lowered a little. It is a matter of ordinary experience that successive draughts of a sweet drink become noticeably less sweet. A phenomenon like successive contrast is seen in taste; following on a sour stimulus distilled water may arouse a sweet sensation. The temperature of the food affects gustation. Food has often more flavour when it is hot; iced food needs much more flavouring. This phenomenon has a practical application in medicine. Thus, for example, fructose used as a test for hepatic function (p. 296) or urea given as a renal function test (p. 607) can be rendered much more agreeable if given thoroughly chilled. Similar considerations apply also to certain drugs, notably tincture of digitalis. This is not true of all drugs, however, as the threshold for quinine is lower when the temperature of the solution is low.

Symmetrical gustatory sweating is the name given to the sweating occurring on the head and face when irritants, such as chillies, are chewed; in a few persons it is aroused by eating chocolate. It is seldom seen in a cool climate or when the body is cooled but is shown by nearly everyone in a tropical climate. This form of sweating is a reflex phenomenon initiated by excitation of the pain receptors in the mouth; the taste receptors are not involved.

REFERENCES

Adrian, E. D. (1956). The action of the mammalian olfactory organ. *J. Laryngol. Otol.* **70**, 1–14.

Baradi, A. F. & Bourne, G. H. (1951). Localization of gustatory and olfactory enzymes in the rabbit and the problems of taste and smell. *Nature,* **168,** 977–979.

Harris, W. (1952). The fifth and seventh cranial nerves in relation to the nervous mechanism of taste sensation. *Brit. med. J.* i, 831–836.

Moncrieff, R. W. (1951). *The Chemical Senses,* 2nd ed. London: Leonard Hill.

Mullins, L. J. (1955). Olfaction. *Am. N.Y. Acad. Sci.* **62,** 247–276.

Schwartz, H. G. & Weddell, G. (1938). Observations on the pathways transmitting the sense of taste. *Brain,* **61,** 99–115.

Stevens, S. S. (Ed.) (1954). *Handbook of Experimental Psychology.* Chap. 29. London: Chapman and Hall.

CHAPTER 39

VISION

STRUCTURE OF THE EYE

A HORIZONTAL cross-section of the right eye to show the general structure is given in Fig. 39, 1. The cornea, lens and vitreous humour are all transparent media with no blood vessels; oxygen and foodstuffs are carried to them by the aqueous humour. The *cornea* also derives part of its oxygen from the atmosphere and if oxygen is excluded the lactic acid concentration rises owing to anaerobic metabolism. A contact lens applied to the cornea may produce sufficient œdema in a few hours to interfere temporarily with vision because of loss of corneal transparency (veiling). This may be due to exclusion of atmospheric oxygen, or interference with the capillary blood supply at the limbus. The cornea is covered by a thin epithelial layer continuous with the conjunctiva at the corneo-sclerotic junction; the substantia propria, the main bulk of the cornea, is formed of criss-crossing layers of collagen and is bounded by elastic laminæ on both front and back; the posterior surface is covered by

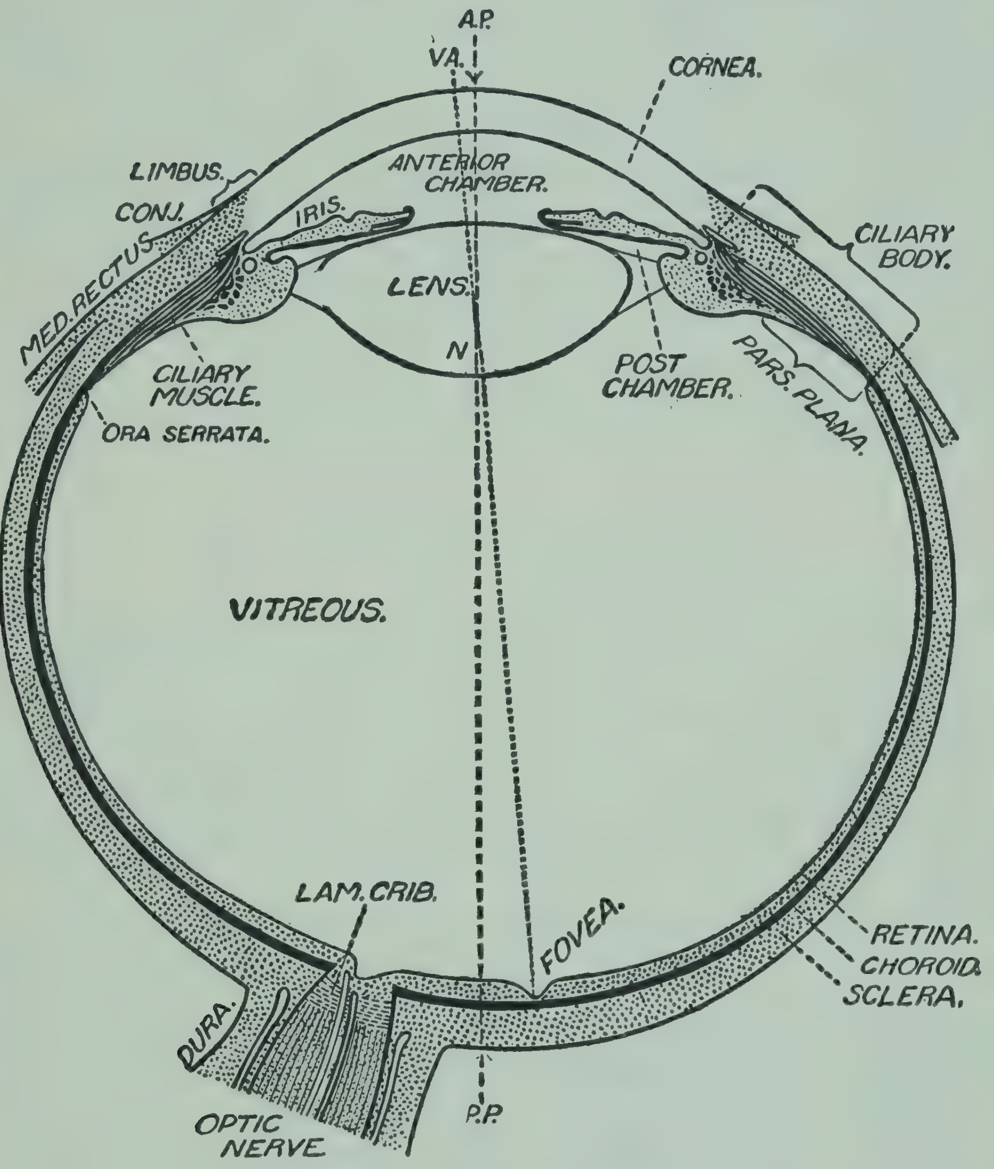

A.P.
V.A.
CORNEA.
LIMBUS.
ANTERIOR CHAMBER.
CONJ.
IRIS.
CILIARY BODY.
MED. RECTUS.
LENS.
N
CILIARY MUSCLE.
POST CHAMBER.
PARS. PLANA.
ORA SERRATA.
VITREOUS.
LAM. CRIB.
FOVEA.
RETINA.
CHOROID.
SCLERA.
DURA.
OPTIC NERVE.
P.P.

Fig. 39, 1

Horizontal section of the eye.

P.P. = posterior pole. A.P. = anterior pole. V.A. = visual axis. (Modified from Salzmann.) (Eugene Wolff (1940). *The Anatomy of the Eye and Orbit,* p. 27. London: Lewis.)

a layer of endothelium. The cornea is richly supplied with free nerve endings. As pain is the only sensation aroused by stimulation of the cornea it has been argued that the function of free nerve endings elsewhere is to signal the sensation of pain. The transparent cornea is continued posteriorly into the opaque white *sclera* which consists of tough fibrous tissue; both sclera and cornea have to withstand the constant tension which is maintained in the eye.

Just inside the sclera is the *choroid*, the vascular pigmented layer, which is continued forwards into the ciliary bodies and then on into the iris. The ciliary bodies produce the aqueous humour and contain the ciliaris muscle which is concerned with accommodation. The stroma of the iris possesses pigment cells and fibrous tissue with radial and circular smooth muscle fibres which control the size of the central aperture or *pupil*. The *retina*, the light sensitive coat of the eye, lies inside the choroid just outside the vitreous; its structure will be described later (p. 681).

The *vitreous humour* is essentially a protein gel and is not hard as its name suggests; its high viscosity is due to the presence of hyaluronic acid. When examined under the ordinary microscope fresh vitreous appears optically empty, but a fine fibrillar scaffolding can be seen with the ultraviolet microscope. If a sample of vitreous is kept for a short time fibrils appear and these are increased by histological fixatives.

PROTECTIVE MECHANISMS

The eye is protected by the strong bony orbit and by the eyelids which are lined by conjunctiva on their inner surfaces. The tarsal glands, which are modified sebaceous glands, excrete a greasy material through ducts opening near the free edge of the lids. Modified sweat glands, ciliary glands, open around the eyelashes which are themselves equipped with sebaceous glands.

The eyelids are closed by the sphincter-like orbicularis oculi supplied by the VIIth (facial) nerve. Reflex closure occurs if the cornea, conjunctiva, or the eyelashes are touched. The afferent arm of the reflex pathway is the ophthalmic division of the trigeminal nerve to its nucleus in the pons; impulses then pass to the superior corpora quadrigemina and finally leave by the facial nerve. Another protective closure reflex is that seen when a bright light is shone on the eye or when an object approaches the eye suddenly; the afferent arm of this reflex is of course the optic nerve. The eyelids are closed for a short time during sneezing and normally they cover the eyes during sleep. The upper eyelid is raised by the levator palpebræ superioris supplied by a branch of the oculomotor nerve. Smooth muscle fibres in the deep part of the levator are supplied by sympathetic fibres derived from the plexus on the carotid artery (see p. 901).

The tear fluid is secreted by the lacrimal glands which lie in the upper and outer parts of the orbits and discharge their secretion through eight to ten ducts to the upper reflexion of the conjunctiva. In addition there are small accessory lacrimal glands which can keep the conjunctiva moist if the main gland is removed. The fluid film on the cornea is renewed by frequent involuntary blinking movements. Since tear fluid does not normally accumulate the amount secreted must be just sufficient to keep pace with the evaporation from the conjunctiva. If, however, foreign bodies or other irritants get on or near the eye, tears are seen. This excessive formation of tear fluid, or lacrimation, also occurs in emotional circumstances or when a bright light is shone into the eye. The excess fluid, if abundant, falls over the edge of the lower lid; smaller amounts may drain by capillary action through minute orifices (puncta lacrimalia), one at the medial margin of each lid, into the lacrimal canaliculi which lead into the lacrimal sac, and the nasolacrimal duct which opens into the nose. The tear fluid is an isotonic aqueous solution of sodium chloride and bicarbonate (pH 7·4) which serves to dilute irritant materials or to wash foreign bodies out of the conjunctival sac. It contains lysozyme, a bactericidal enzyme, which reduces the bacterial content of the conjunctival sac.

Blinking has, as mentioned above, the effect of protecting the eye and of renewing the fluid film over it. Although we are usually completely unaware of it, we are 'blacked-out' for about 0·3 sec during each blink. The rate of blinking varies widely in different persons within the range 5 to 50 per minute and is about 20 per minute on the average. If the interblink period is 3 seconds and the black-out is 0·3 second we are blacked-out for 10 per cent. of our waking time. Thus blinking may put an upper limit to the efficiency of observation, especially of events of short duration.

Aqueous Humour

The aqueous humour carries nutrients to the non-vascular structures of the eye. It contains sodium chloride, small amounts of protein and a reducing sugar, has a high oxygen tension and about the same osmotic pressure as blood. The fluid is produced under pressure so that the normal *intra-ocular tension* in man is 25 to 30 mm Hg. Experiments on animals show that the intra-ocular pressure remains fairly constant even when the arterial pressure fluctuates widely although it rises a little when the external ocular muscles contract and on winking. As will be seen later, the focusing mechanism of the eye depends upon the presence of a fairly constant intra-ocular tension. An excessive reduction in tension slackens the suspensory ligaments of the lens and allows the latter to bulge; conversely if the tension is too high, as in glaucoma, the ciliaris muscles may not be able to bring about accommodation. Restriction of the retinal circulation, with resultant damage to the retina, may also be caused by too high an intra-ocular pressure.

Since its composition is much like plasma minus proteins, the aqueous humour has been regarded simply as a filtrate from the blood plasma in the capillaries of the ciliary body and iris. To maintain a flow of blood in the intra-ocular veins the pressure in them must be above the intra-ocular pressure; the pressure in the capillaries must be still higher, say 50 mm Hg, and this is sufficient to account for filtration. Recent more exact work shows, however, that the composition of the aqueous is considerably different from that of a simple filtrate of plasma. In the dog and cat, for example, the concentrations of Na^+ and Cl^- in the aqueous are much higher and those of urea and glucose are lower than would be found in a simple filtrate from plasma. The low concentration of glucose can be explained by its utilization by surrounding tissues, especially the lens, and in conformity with this the concentration of glucose in the vitreous is only half that in the aqueous; but this cannot be the explanation in the case of urea. The pathway for the production of aqueous fluid must be transcellular because of the high selectivity of the blood-aqueous barrier. For example, glucose passes rapidly from blood to aqueous whereas sucrose passes only slowly; but glycine with a molecular weight less than half that of glucose passes more slowly than glucose.

The aqueous humour, vitreous body, lens and cornea all contain large amounts of ascorbic acid (vitamin C) which may play a part in the glucose metabolism of the avascular parts of the eye. However, even in severe scurvy (p. 140) no eye changes are observed.

Since the intra-ocular tension remains constant, it is obvious that the production of aqueous humour must be balanced by an equal loss from the eye. The possible channels for the escape of the fluid are shown in Fig. 39, 2. It was suggested more than fifty years ago that the main route of escape is via the filtration angle into the sinus venosus scleræ and into the ciliary veins. Stratification of the contents of these collecting veins has been noted owing to the presence side by side of streams of blood and aqueous humour. In man Indian ink injected into the anterior chamber just before enucleation of the eye was found to follow the same route. If the filtration angle (Fig. 39, 2) is blocked by debris after the operation of needling the lens the intra-ocular pressure rises. The sinus venosus scleræ being a region of low blood pressure is almost certainly the main drainage system. By introducing inulin or *p*-amino-hippuric acid into the anterior chamber of the eye of an animal and measuring the fall in concentration with time it is found that about 1 per cent. of the volume of the aqueous humour drains away per minute. The drainage of the aqueous humour, unlike its production, is an unspecific mechanical process.

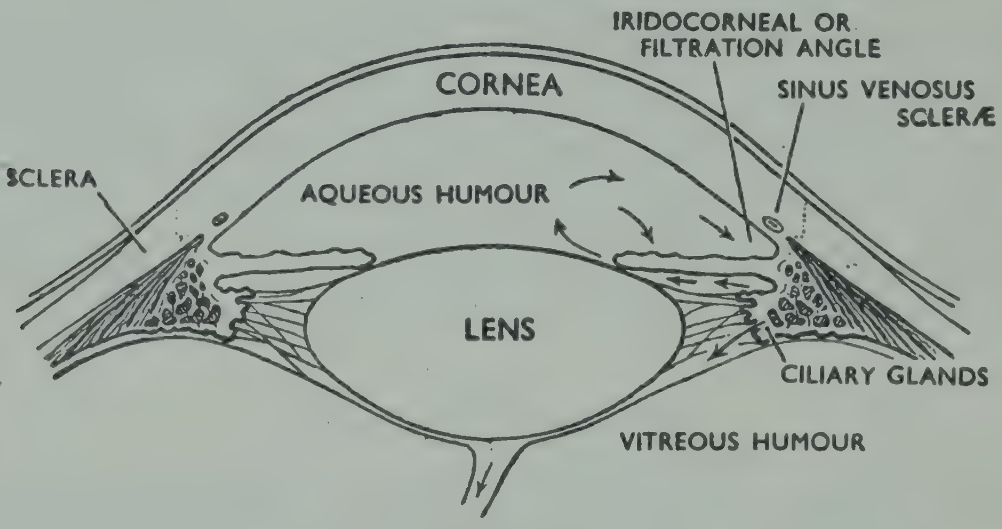


FIG. 39, 2

To illustrate the probable source of aqueous humour and its routes of absorption into the circulation.

OPTICS OF THE EYE

All the transparent structures of the eye play a part in forming images of external objects on the retina which is the light-sensitive layer of the eye. The refractive index of all the transparent media is 1·33 with the exception of the lens which, owing to its peculiar construction, has an equivalent refractive index of 1·42.

Microscopic preparations show that the lens is composed of ribbon-like fibres, modified epithelial cells of ectodermal origin, arranged in concentric laminæ which are relatively soft in the cortex and more dense in the nucleus of the lens. Thus the central nucleus has a higher refractive index than the peripheral laminæ. The lens is enclosed in a capsule which is attached to the ciliary bodies by the suspensory ligaments; when these are slackened off, or when the lens is removed from the eye, the lens bulges by its own elasticity. It contains about 65 per cent. of water and 30 per cent. of protein; the water content increases with age. If the lens is excised and placed in a respirometer it continues to take up oxygen and produce carbon dioxide. Since it degenerates if the oxygen supply becomes inadequate there is no doubt that the lens is a living tissue in spite of its lack of blood supply. It contains glycolytic enzymes which break down glucose to lactic acid. Glutathione seems to play an important part in the oxidative processes; in senile cataract (opacity) there is no glutathione in the lens.

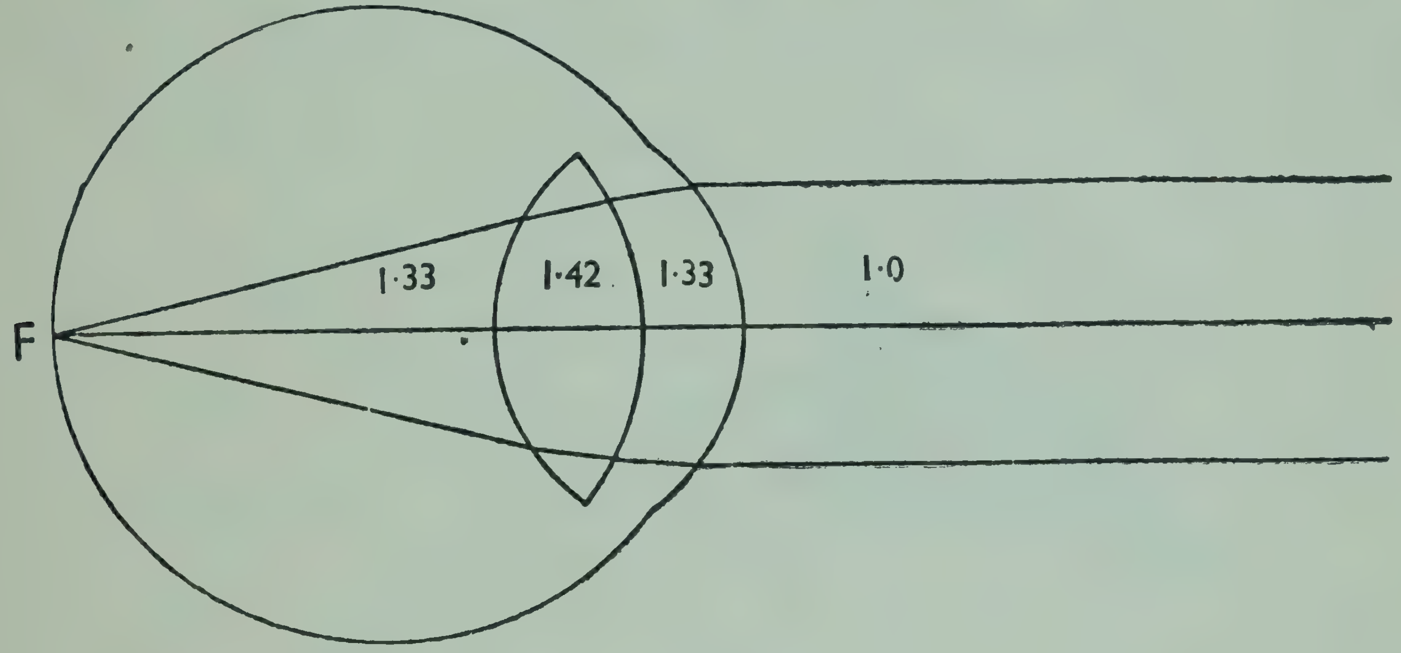


FIG. 39, 3

Emmetropic eye. This diagram shows how the rays of light are refracted where there is a difference of refractive index between any two media. The figures give the approximate refractive indices of the media.

In a lens system light is refracted or bent out of its original direction where there is a change of refractive index. A ray arriving at the air-cornea interface is refracted there and passes on straight through the aqueous (since the refractive indices of cornea and aqueous are practically the same) to the surface of the lens, where it is again bent towards the normal. On emerging from the lens it is bent away from the normal to the lens surface since the vitreous has a lower refractive index (1·33) than the lens. The greatest amount of refraction occurs at the air-cornea surface (Fig. 39, 3). The curvature of the cornea can be assessed by placing an object of known size in front of it and measuring the size of the virtual image formed by the anterior convex surface. If a series of concentric rings (Placido's disc) is viewed by reflexion from the cornea the shape of the rings in the image (normally circular) reveals corneal distortions. The focal length of the cornea measured in this way is about 24 mm or 42 dioptres. The focal length of the whole eye, that is the distance from the posterior node (N in Fig. 39, 14) to the retina, is 17 mm or 59 dioptres. (The power of a lens in dioptres is given by the reciprocal of its focal length in metres.) The dioptric power of the eye after removal of the lens is about 43 dioptres; the lens thus accounts for only about 16 dioptres of the total refractive power of the eye. If the eyes are open during swimming under water the refractive power of the cornea is lost, because the refractive indices of water and cornea are nearly the same; the eye virtually becomes long-sighted and distinct vision is not possible. If a patient becomes blind because of an opacity of his lens (a cataract), the lens can be removed and he is then supplied with a spectacle lens of about 10 dioptres for ordinary vision and 14 dioptres for reading. This is less than the actual power of the eye lens but, since the spectacle lens is further away from the retina than the eye lens was, less convergence is needed and a smaller power is adequate.

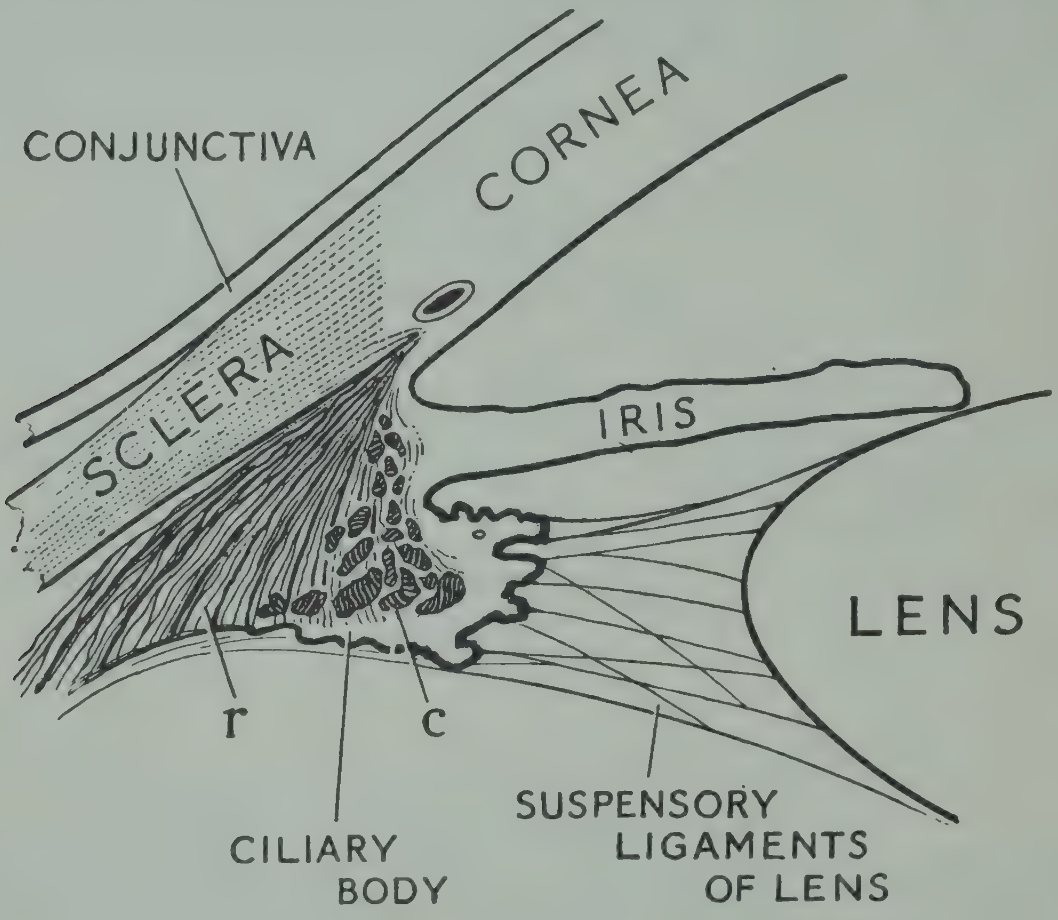


Fig. 39, 4

Section through the cornea-sclerotic junction to show, *r*, the radial muscle fibres and, *c*, the circular muscle fibres of the ciliary body cut transversely.

Accommodation

A fixed lens system can provide clear images of objects situated at one distance only. Accommodation or alteration of the refractive power of the lens of the human eye is brought about by the ciliaris muscle. This consists of two sets of fibres as indicated in Fig. 39, 4. The superficial or radial fibres arise near the cornea-sclerotic junction and pass backwards to the ciliary processes. The circular fibres form a sphincter muscle in the deeper part of the ciliary body. The muscle fibres are however somewhat irregularly arranged and their separation into radial and circular fibres may be artificial.

When the eye is at rest the intra-ocular tension keeps the ciliary processes out and back, and so puts tension on the suspensory ligament of the lens which is thus kept somewhat flattened. In this condition images of distant objects are in focus. When the two sets of fibres of the ciliaris muscle contract the ciliary processes are drawn forwards and are made to occupy a smaller circle; the suspensory ligaments become slack and the lens bulges, so that near objects are brought to a focus on the retina. By observing the images (Purkinje-Sanson images), formed at the anterior and posterior surfaces of the lens, of lamps placed near the eye it can be shown that the posterior surface of the lens is little altered in accommodation, whereas the anterior surface becomes more curved and moves forward slightly (Fig. 39, 10).

When the gaze is moved from a distant to a near object the retinal images are blurred and disparate, that is they do not fall on corresponding points on the retinæ (see p. 700). Accommodation abolishes the blurring by bringing the divergent rays of light to a focus on the retina and fusion of the disparate images is brought about by convergence. These corrections occur so quickly that the blurring is not normally noticed. These adjustments, especially the convergence adjustment, may be regarded as voluntary acts dependent on knowledge of the position of the near object. With emmetropic eyes in perfect muscle balance when the eyes look from infinity to an object 1 m distant they accommodate 1 D; when they converge to 0·5 m, the accommodation is 2 D and so on until the near point is reached. There is thus a linkage between accommodation and convergence.

In monocular vision other factors come into action as shown by Fincham. When a weak negative lens which diverges the rays is put in front of a normal eye (the other being covered) gazing at a distant object, accommodation occurs reflexly and the distant object remains in focus; removal of the lens is followed by a relaxation of accommodation. Convergence of light rays from a distant object brought about by putting a convex lens in front of the eye never produces positive accommodation but may produce a relaxation in hypermetropes. In such monocular experiments, as the change in accommodation occurs promptly and in the correct direction, it is unlikely that it is the blurring of the retinal image which starts off the reflex but more likely that it is the divergence or convergence of the light. The vergence may be detected as a result of the normal chromatic aberration of the eye. The retinal image of a white or coloured object has fringes which vary in colour as the refractive power of the eye is changed and these colour changes could presumably indicate in which direction accommodation should be adjusted. Many subjects show no reaction to changes in vergence of monochromatic light. In others change of vergence of monochromatic light does produce change in accommodation, perhaps as a result of scanning movements of the eye.

The afferent pathway for accommodation is the visual pathway to the calcarine cortex and we must assume that the information travels to the superior corpus quadrigeminum and to the Edinger-Westphal nucleus in the mid brain. The efferent part of the arc is formed by the parasympathetic fibres of the third nerve arising in the Edinger-Westphal nucleus which relay in the ciliary ganglion and pass in the short ciliary nerves to the ciliaris muscle. Stimulation of this nerve makes the ciliaris muscle contract and produces an increase in the refractive power of the eye; eserine placed in the conjunctival sac travels by diffusion to the ciliary body and, by its anticholinesterase activity, produces a spasm of the ciliaris muscle. Atropine which can travel by the same route paralyses the muscle so that accommodation cannot occur and near objects cannot be seen. The ciliaris muscle is antagonized by the elasticity of the lens capsule; thus when the muscle relaxes a little, the lens flattens a little and a nice balance is reached. This may in fact apply to man but in the cat, stimulation of the sympathetic reduces the refractive power of the eye. Photographic records of the Purkinje-Sanson images show that the time for accommodation from far to near is about half a second.

Range of Accommodation.—In the emmetropic or normal eye the range is measured by finding the nearest point to the eye at which distinct vision is possible; the near point gradually recedes as age advances, as shown in Fig. 39, 5. The best method of finding the amplitude of accommodation is to find the range of plus and minus (convex and concave) lenses that can be put in front of the eye without causing the image of an external object to appear blurred; the amplitude diminishes with age (as shown in Fig. 39, 6) due to the hardening and flattening of the lens and not to a failure of the ciliaris muscle. This perfectly normal condition is termed *presbyopia*. It only begins to be obvious about 40 years of age when near work like threading a needle or putting a cannula into a vein is being done. Later, owing to the greater recession of the near point, newspapers have to be held so far away from the eye that reading of fine print becomes impossible. The elderly person with myopia (p. 679) has the advantage that he can read print near at hand if he takes his glasses off.

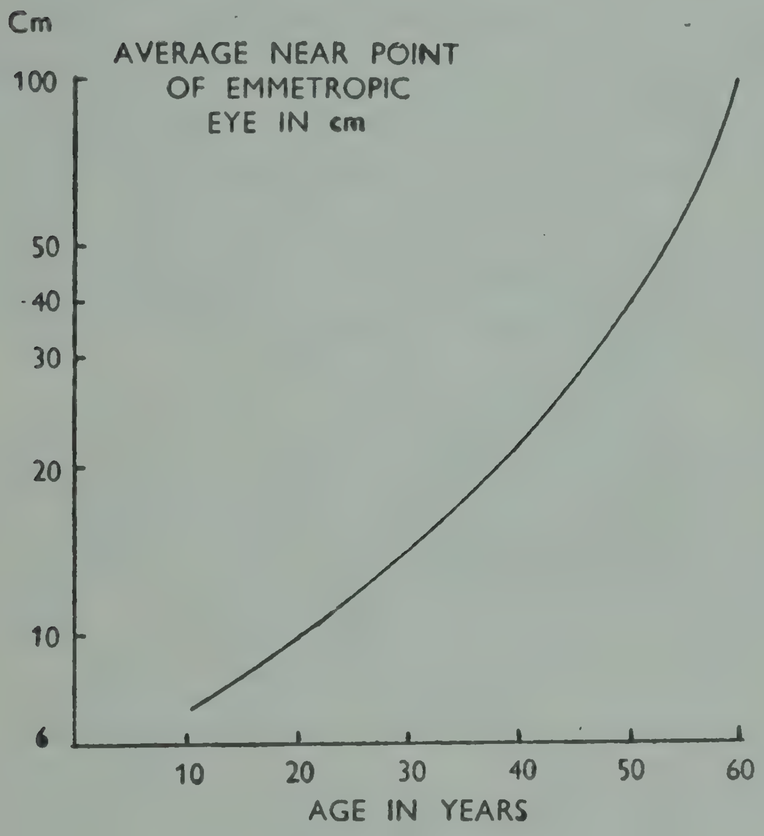


FIG. 39, 5

To show how the near point varies with age.

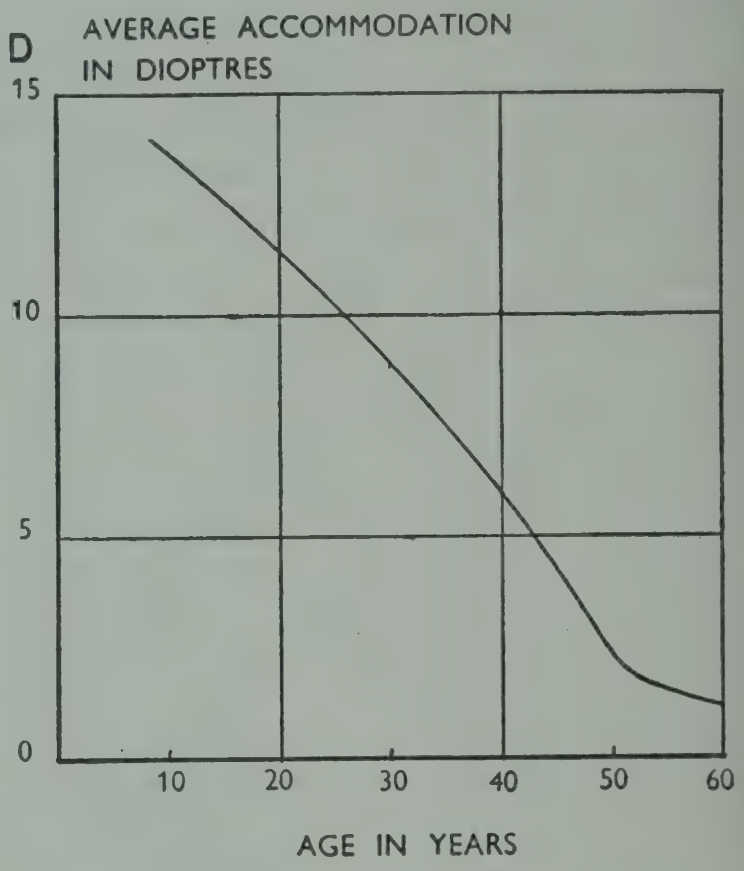


FIG. 39, 6

To show how the amplitude of accommodation varies with age.

THE IRIS

Innervation of the Iris

The iris, being embryologically the continuation forwards of the choroid coat, contains much pigment and is opaque. The aperture or pupil is controlled by the radial and circular muscle fibres of the iris. The radial fibres forming the dilatator pupillæ are innervated in the following way.

Fibres from the hypothalamus pass down the spinal cord and terminate in the ciliospinal centre of Budge; second order neurones then emerge from eighth cervical segment and the upper three or four thoracic segments as white rami communicantes and pass up the sympathetic chain to the superior cervical ganglion. There postganglionic fibres arise and pass along the coats of the internal carotid artery to the ophthalmic division of the trigeminal nerve and long ciliary nerves to the eye. These are adrenergic fibres. The sphincter muscle of the iris is innervated from the third nerve nucleus (Edinger-Westphal nucleus) by fibres running in the third nerve to the ciliary ganglion where there is a synapse; impulses are then carried in medullated postganglionic fibres in the short ciliary nerves. The transmitter is acetylcholine. The third nerve nucleus receives inhibitory fibres from the sympathetic centre in the hypothalamus mentioned above. A diagram of the innervation is given in Fig. 49, 14.

The size of the pupil is controlled mainly by the variations in tone of the sphincter muscle, that is, by the parasympathetic supply; the sympathetic plays a minor role and provides only a fairly steady tone to facilitate the results of relaxation of the sphincter. This is well illustrated by the finding that after cutting the sympathetic on one side the pupillary reflexes are normal; the only difference is that the pupil on the affected side is smaller. A further illustration of the dominance of the sphincter muscle is that stimulation of a certain area of the frontal cortex is followed by dilatation of the pupil, that is, by relaxation of the sphincter, even when the sympathetic is cut. This finding could also be interpreted as evidence of reciprocal innervation of the two muscles of the iris (compare p. 845).

Stimulation of the sympathetic nerves to the eye leads to dilatation of the pupil, retraction of the upper lid because of the contraction of the unstriped muscle in the deep part of the levator palpebræ superioris and protrusion of the eyeball (proptosis). Injuries of the cervical cord in man, if the sympathetic pathway is interrupted, are accompanied by a small pupil (miosis), a ptosis (falling) of the upper eyelid and an apparent retraction (enopthalmos) of the eyeball (Horner's syndrome).

Functions of the Iris

These are much the same as those of the iris diaphragm of a camera. A small pupil prevents light passing through the more peripheral parts of the lens system, cuts down the amount of light reaching the retina and reduces the effects of chromatic and spherical aberration. A small pupil also increases the depth of field, that is, the range within which objects are clearly seen without altering the accommodation.

The pupil contracts when light falls on the eye. This is the *light reflex*. Impulses pass from the retinal receptors to fibres in the optic tract and travel without decussation to the pretectal area immediately in front of the superior quadrigeminal body (Fig. 49, 13). Impulses then pass to the Edinger-Westphal nuclei of the third cranial nerves on both sides, thus accounting for the fact that when light shines into one eye not only does the pupil of that eye contract (direct light reflex) but also the pupil of the other eye (consensual light reflex).

In certain diseases of the nervous system the light reflex is abolished but the pupil still contracts on accommodation. The lesion is probably in the pretectal area; in animals destruction of the superior quadrigeminal body alone does not abolish the light reflex but simultaneous destruction of the superior quadrigeminal body and the pretectal area does.

Although the retina itself responds quickly to light, the pupil begins to contract only after a latent period of about 0·5 second and reaches its minimum size in five seconds or so. This slow response might be expected, since the sphincter pupillæ is composed of smooth muscle, but this muscle can, in the condition known as hippus, show rapid contractions and relaxations. There must, therefore, be some sort of delay mechanism in the reflex pathway.

The amount of the contraction of the pupil on exposure to light depends on the previous lighting conditions and the rate at which the light increases. For example, there may be little alteration in the size of the pupil throughout the day although the illumination varies over a considerable range. If, however, a person goes quickly from the dark into the light the pupils contract very quickly but, during the succeeding ten minutes, as the retina becomes adapted, the pupils gradually become larger again although the intensity of the light remains constant.

In sleep the pupils contract; if the eyelids are closed in the waking state the pupils dilate. When accommodation occurs and the visual axes converge the pupils contract; this is described as the near reflex (Fig. 39, 10). The advantage of this association is that the depth of field is increased, thus diminishing somewhat the necessity for rapid changes of accommodation during reading. Stimulation of the cornea and the conjunctiva of the lids produces a contraction of both pupils. Some drugs, such as morphine and eserine, produce very small pupils (pin-point pupils). In old age the pupils are smaller and less active.

Dilatation of the pupils occurs when many of the situations just described are reversed, for example, when the transition is from light to dark or from near vision to distant vision. Stimulation of any sensory nerve (except the trigeminal) causes a dilatation of the pupil presumably because the afferent impulses reach the cells in the cord which give rise to the sympathetic fibres and cause a dilatation of the pupil. Stimulation of the central end of the splanchnic nerve causes dilatation of the pupil and this reflex has been used to trace the afferent fibres from the intestine. In conditions such as fear, pain, or asphyxia, which call out adrenaline from the adrenals, the pupils dilate as in stimulation of the sympathetic nerves to the eye. Certain drugs, including atropine, homatropine, cocaine, adrenaline, produce a dilatation; alcohol and chloroform poisoning also have this effect.

OPTICAL PROPERTIES OF THE EYE

In the design of a camera lens two of the difficulties to be overcome are *spherical* and *chromatic aberration.* The former is due to the greater refractive power of the peripheral parts of a homogeneous spherical lens which brings the rays passing through the periphery to a focus in front of those passing through the centre. This defect is largely overcome in the human eye by the high refractive index of the central part of the lens. Chromatic aberration is due to the nature of light, so that rays of short wavelength are bent more than rays of long wavelength when they pass from a medium of low refractive index to one of a higher index. In the case of the eye, when the red rays are sharply focused on the retina, the blue rays come to a focus nearly 0·4 mm in front of the retina. The eye is normally focused so that the brightest rays come to a focus on the retina while the rays of longer and shorter wavelengths form blurred discs of relatively low intensity surrounding the bright centre, which is usually yellow. While the images at the fovea centralis are sharp the images on the peripheral parts of the retina are less distinct, but this is no great disadvantage as the function of the peripheral retina is chiefly to call attention to objects in the visual field. Once an object is noticed it can be examined in detail by the fovea. Although the eye exhibits chromatic aberration this defect is of no importance since, if it is corrected by a suitable lens or eliminated by using monochromatic light, visual acuity is improved to an insignificant extent.

If the eye is to function satisfactorily there must be a balance between the length of the eye and its refractive power. Quite often, however, the eye is too long, either because of defective nutrition or for some genetic reason, and short-sight or *myopia* results. If a child showing this defect is allowed to read small print brought close to the eye the condition may become progressive and myopia should, therefore, be treated by supplying suitable lenses and by prohibiting too much near work. The condition is called short-sight because only objects near the eye can be seen clearly; since rays from points at a distance are brought to a focus in front of the retina, distant objects cannot be seen clearly (Fig. 39, 7, D). The refractive error can be corrected by concave spectacle lenses (Fig. 39, 7, E). Normally on looking at a near object convergence of the optical axes of the eyes and accommodation occur together, but this association is disturbed in myopia and consequently eyestrain and divergence of the visual axes (external squint) are apt to occur.

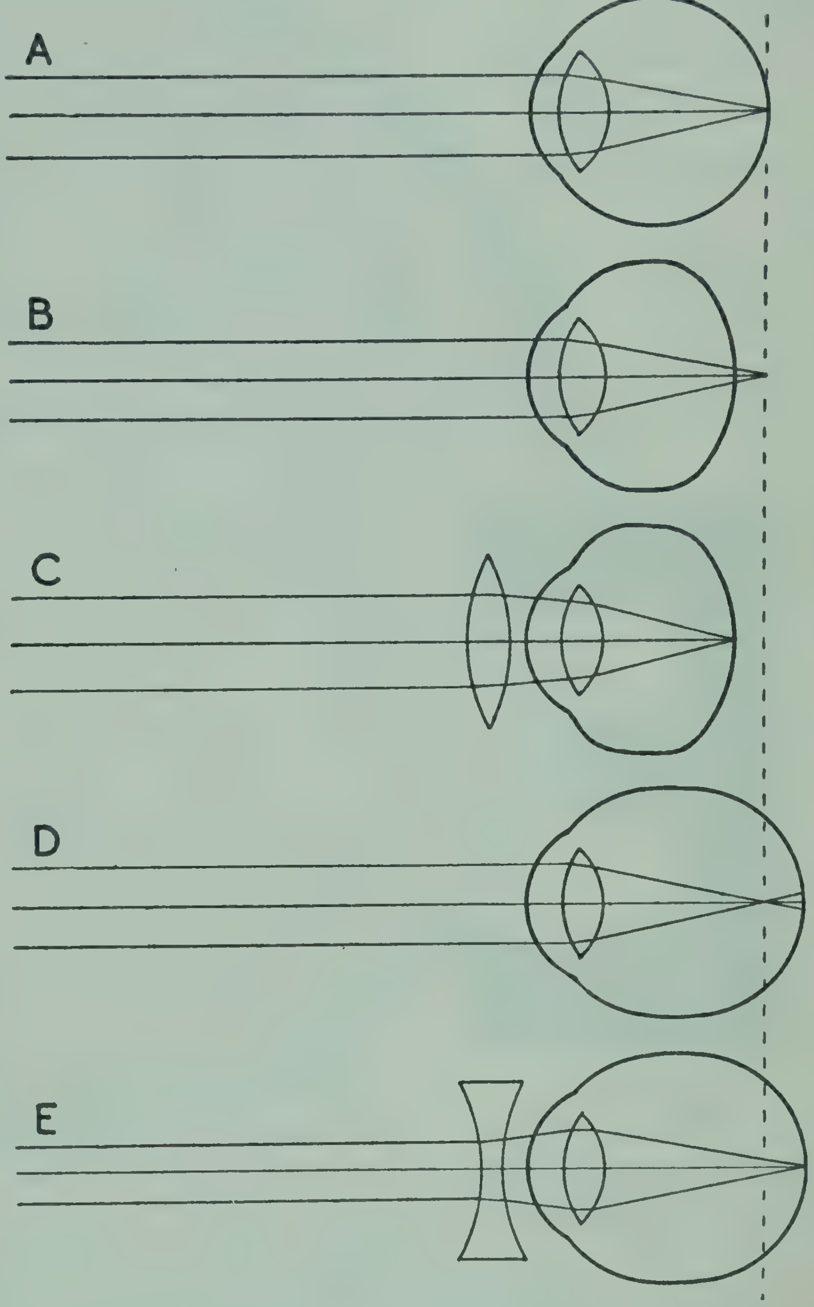


FIG. 39, 7

Diagram illustrating the common refractive errors of the eye and the methods of correction. A, emmetropic eye. B and C, hypermetropic eye. D and E, myopic eye.

In some children the long-sighted or *hypermetropic* condition persists because the eye remains too short. In older people hypermetropia may result from a reduction in the refractive power of the lens. In both cases rays of light from distant objects are brought to a focus behind the retina (Fig. 39, 7, B) when the ciliaris muscle is relaxed. The ciliaris muscles are, however, brought into action by the accommodation reflex and so distant objects are seen clearly. Obviously, if accommodation is used to enable the subject to see distant objects, the nearest point of distinct vision may be well away from the eye and near objects cannot be seen clearly. Eyestrain, fatigue and headaches are apt to result from the continued effort of accommodation. Since accommodation is required without the necessity for convergence the dissociation may result in an internal squint, that is the optical axes of the eyes are convergent at rest. As the eye is too short, the filtration angle is rather narrow and there is a tendency for it to become blocked; glaucoma occurs more than twice as frequently in hypermetropes as compared with myopes. The refractive error can be corrected by a suitable convex (plus) lens (Fig. 39, 7, C).

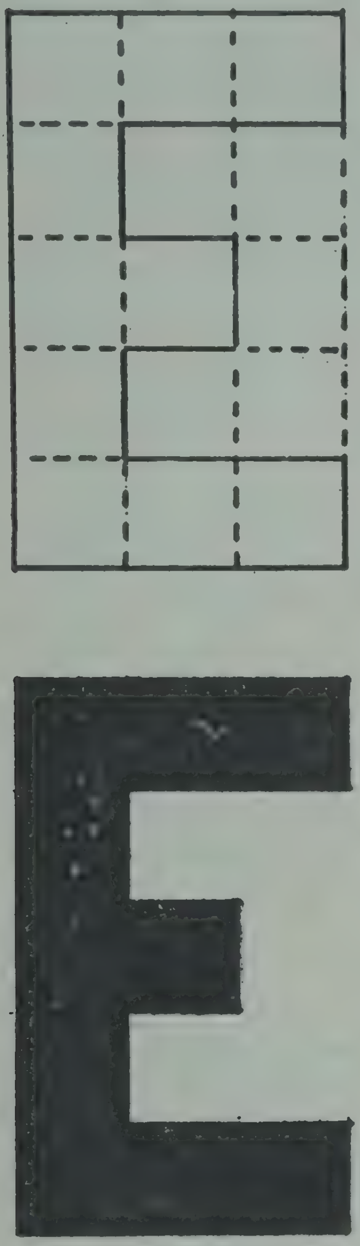


Fig. 39, 8

The method of construction of Snellen's types.

The classical explanation, as given in the preceding paragraphs, that myopia and hypermetropia are due to departures from the normal axial length of the eye, assumes that the cornea-lens system is constant. This is not so. Measurements of a large number of eyes show that the refractive power of the cornea varies from 39 to 47 D and that of the lens from 14 to 25 D. X-ray measurements of the axial length of emmetropic eyes show that it varies from 22 to 26 mm; further, although hypermetropic eyes tend to have a shorter axis and myopic a longer, it is only in the severe degrees of these defects that the axial length falls outside the normal limits. The distribution of refractive errors in a population shows that the three variables, cornea, lens and axial length, are not associated in a random fashion but must be correlated to produce the excess of emmetropes. There must be a correlation mechanism to maintain something like emmetropia during the period when the new-born eye, 16 mm long, grows to the adult length of 24 mm. In any case myopia usually declares itself during adolescence when the emmetropic eye is established and stable, there being little growth in length after three years.

It often happens that different meridians (lines radiating from the centre of the cornea of the eye) have different focal lengths so that clear images of objects cannot be obtained. If the meridian with the shortest focal length is at right angles to the meridian with the longest, and there is a gradual transition between these meridians, the condition is described as regular *astigmatism.* If a subject with regular astigmatism looks at a fan of radiating lines, some lines appear clear and distinct while others at right angles are blurred. This error can be corrected by supplying the patient with a cylindrical lens of the correct power with its axis at such a position that the focal lengths of the maximum and minimum meridians are made equal. If a spherical correction is also required, it is usually ground on one side of the spectacle lens with the cylindrical correction on the other.

If the cornea acquires an irregular surface, due for example to a healed ulceration, irregular astigmatism is present. It is obvious that this cannot be corrected by a spectacle lens, but a contact lens fitted over the cornea may improve vision considerably by eliminating the optical effect of the anterior corneal surface and replacing it with the regular front surface of the contact lens.

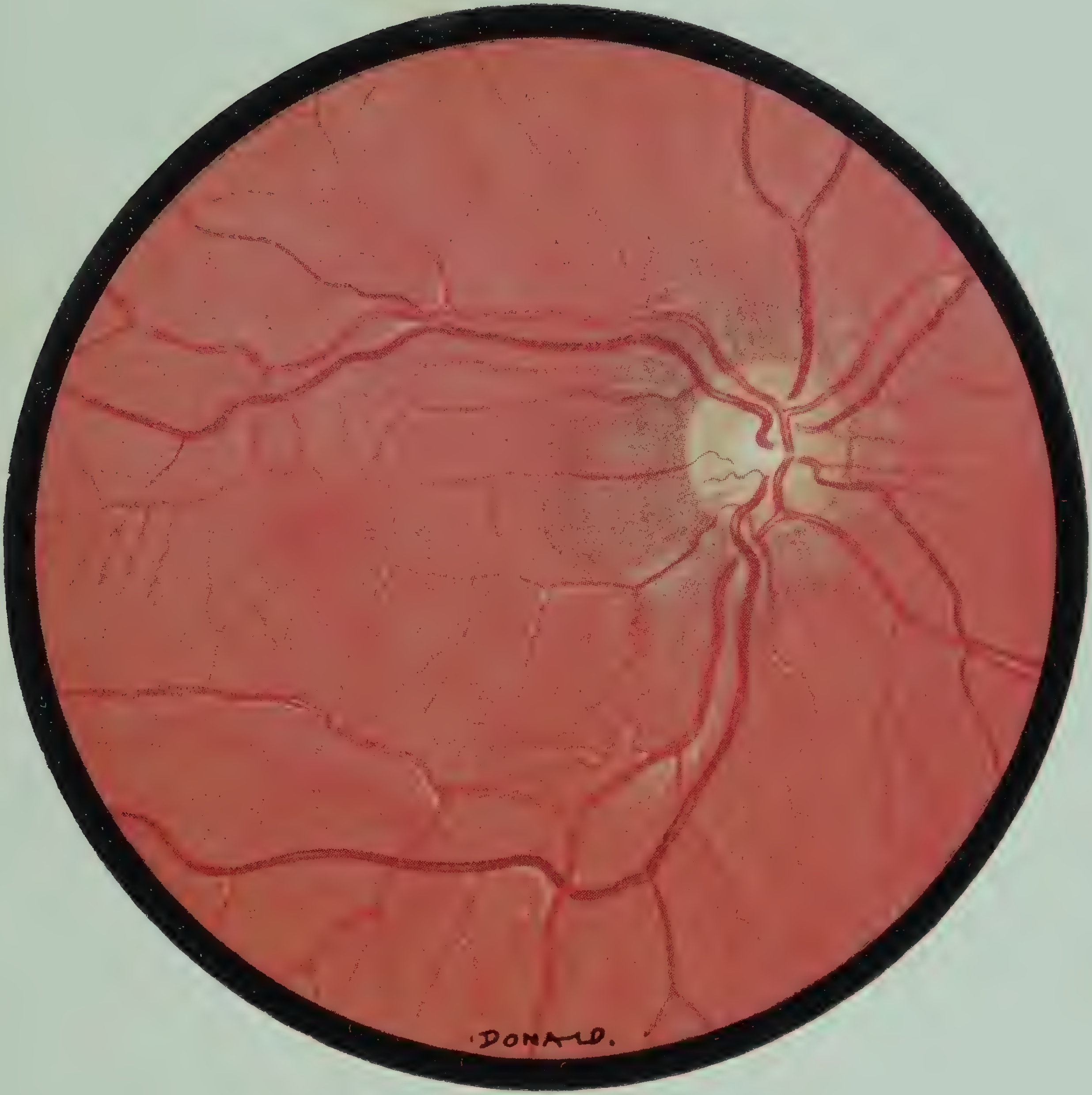


FIG. 39, 9

The fundus of the normal human right eye as seen through an ophthalmoscope. This view of the interior of the eye shows on the right the yellowish optic disc with the retinal vessels emerging from its centre. The macula is seen as a slightly darker area on the left of the figure. Note that minute vessels converge upon it but do not cross it. The thin dark vessels are veins and the narrower lighter vessels are arteries. The pink background is due to blood in the choroid seen through the transparent retina.

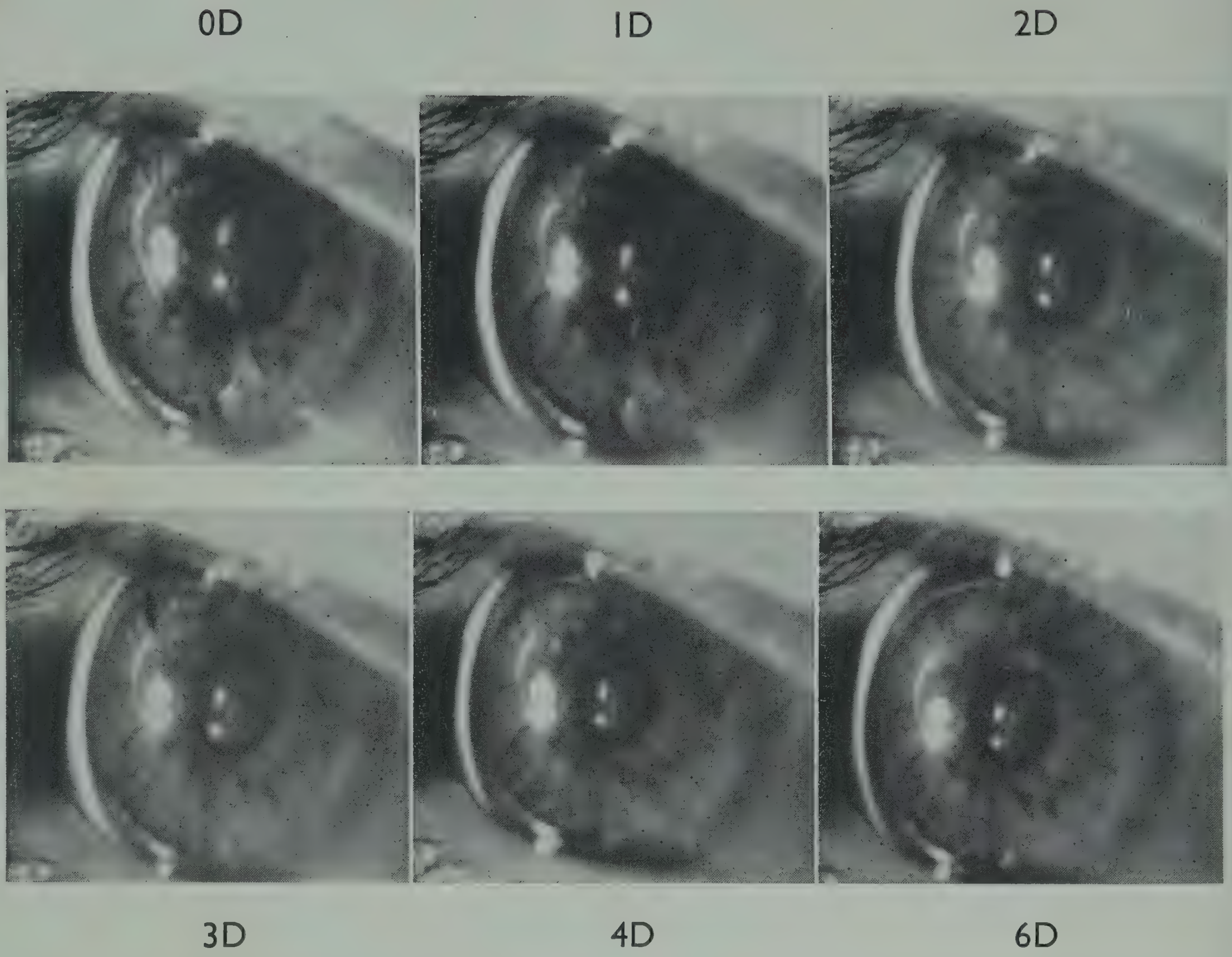


Fig. 39, 10

A series of photographs of an eye kept in the dark except for the flashes of light needed to take the photograph. The source of light was doubled and is reflected from the anterior surface of the cornea and from the anterior surface of the lens (the two smaller spots near the centre of the pupil). The accommodation was produced by placing a dim fixation point at suitable distances from the eye. Note that, as accommodation increases, the two images from the front surface of the lens become smaller, brighter and closer together; this indicates that this surface becomes more convex. The pupil also becomes smaller as accommodation increases. The amount of accommodation is indicated in dioptres (0D to 6D) next each photograph. (F. W. Campbell & J. A. E. Primrose (1953), *Trans. Ophthal.* Soc. **73,** 356).

METHODS OF MEASURING REFRACTIVE ERRORS

Persons with normal sight can distinguish black bars against a well-illuminated white background if they subtend an angle of one minute at the eye. Visual acuity is usually measured by asking the subject to read Snellen's test types, which are constructed in rectangles, as shown in Fig. 39, 8. Each line of letters is marked with the distance in metres at which the sides of the small squares subtend an angle of one minute. A normal subject at 6 metres from the test types should see the types marked 6 and his vision is reported as 6/6. If a subject with defective vision standing 6 metres from the chart can read only the line marked 60 his vision is reported as 6/60. Defective vision may be due to defects in the retina or central nervous system or in the refractive media of the eye. If the subject can see brightly lit test type clearly when he looks through a very small aperture which makes the eye into a 'pinhole camera' the cause of the defective vision must be a refractive error. The amount of the refractive error can be discovered by placing lenses of various powers, convex or concave, in front of the eyes until the best visual acuity is obtained. Where difficulties of age or language prevent the co-operation of the patient the refractive error can be estimated by retinoscopy and other methods which are too technical to be described here.

THE RETINA

STRUCTURE OF THE RETINA

The retina is formed by the optic vesicle which grows out from the fore-brain of the embryo (Fig. 39, 11). The outer wall of the vesicle forms the pigmented layer of the retina, while all other layers are derived from the inner wall. If the ectoderm is followed into the fore-brain and into the optic vesicle it will be found that the skin surface of the embryo

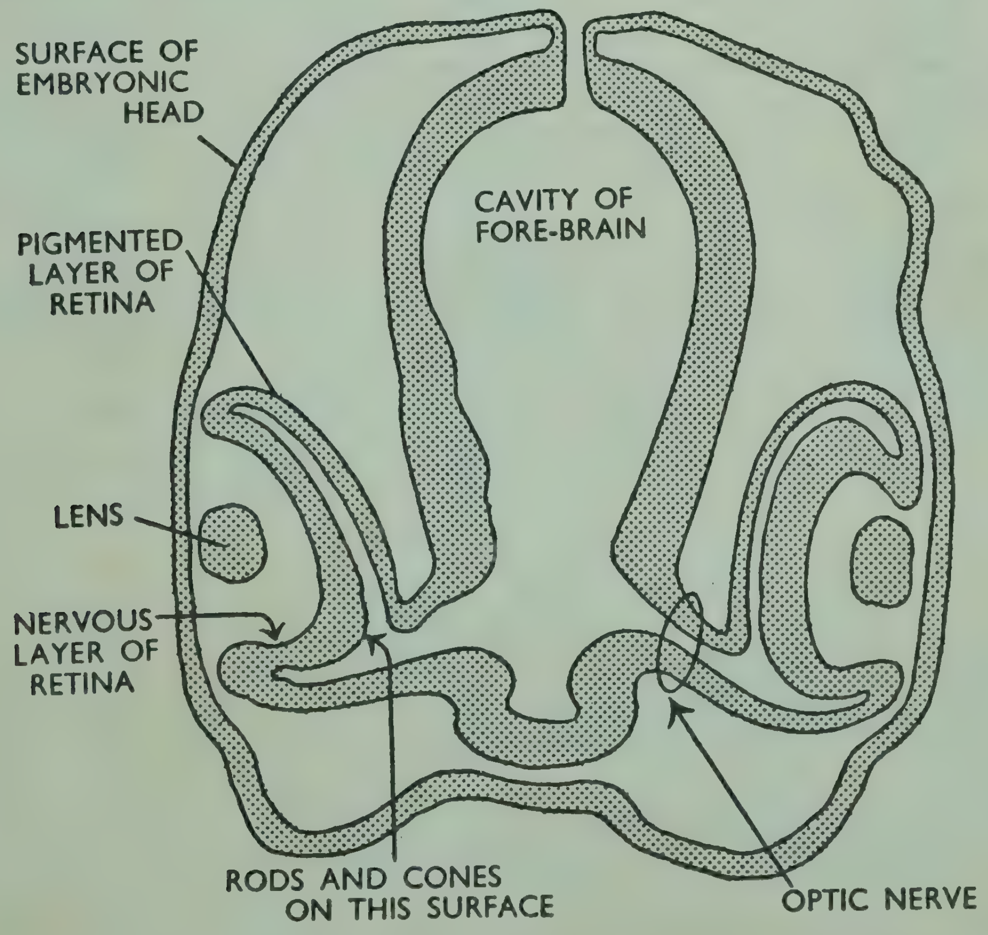

SURFACE OF EMBRYONIC HEAD
PIGMENTED LAYER OF RETINA
LENS
NERVOUS LAYER OF RETINA
CAVITY OF FORE-BRAIN
RODS AND CONES ON THIS SURFACE
OPTIC NERVE

FIG. 39, 11

Schematic section of embryonic head to show how the eye is developed in relation to the brain. (*After* Gray's *Anatomy*.)

corresponds to the layer of cells lining the cavity of the vesicle (Fig. 39, 11). This is the layer which contains the sensory elements, the rods and cones. The structure of the retina is shown diagrammatically in Fig. 39, 12. The retina passes forwards as far as the ora serrata (Fig. 39, 1) at the equator of the eyeball; beyond this the nervous layer, represented by a layer of columnar cells with no sensory function, passes forward together with the pigmented layer to the posterior surface of the iris. Near the centre of the posterior part of the retina opposite the pupil is an oval yellowish area, the *macula lutea*, with a central depression, the *fovea centralis* (Fig. 39, 1). The point of exit of the optic nerve is shown by the whitish optic disc about 1·5 mm in diameter situated about 3 mm to the nasal side of the fovea centralis. The central artery and vein of the retina pierce the optic disc to supply the retina. Since, in spite of their proximity, there is no anastomosis between the retinal and choroidal vessels, blockage of the retinal artery is followed by sudden and complete blindness.

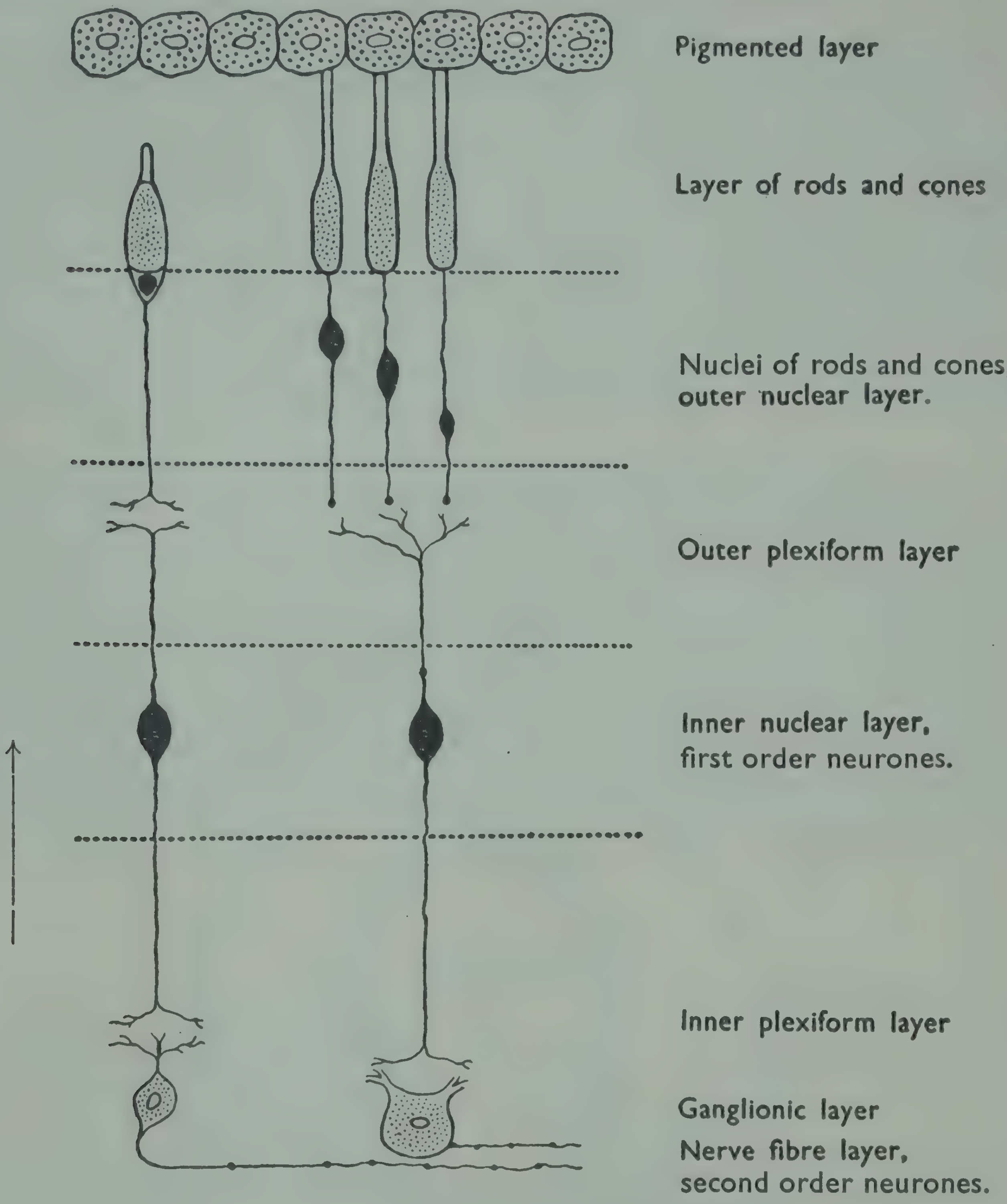

Pigmented layer
Layer of rods and cones
Nuclei of rods and cones, outer nuclear layer.
Outer plexiform layer
Inner nuclear layer, first order neurones.
Inner plexiform layer
Ganglionic layer
Nerve fibre layer, second order neurones.

FIG. 39, 12

Simple scheme of connexions between the rods and cones and the optic nerve fibres. In addition there are horizontal cells and also neuroglia cells as in other parts of the brain. Second order neurones become myelinated after they pass out of the eyeball. Light passes through the retina to the rods and cones in the direction of the arrow.

The structures inside the eye can be observed in life by means of the *ophthalmoscope.* A beam of light from an electric lamp is directed through the subject's pupil through the transparent media on to the fundus oculi.

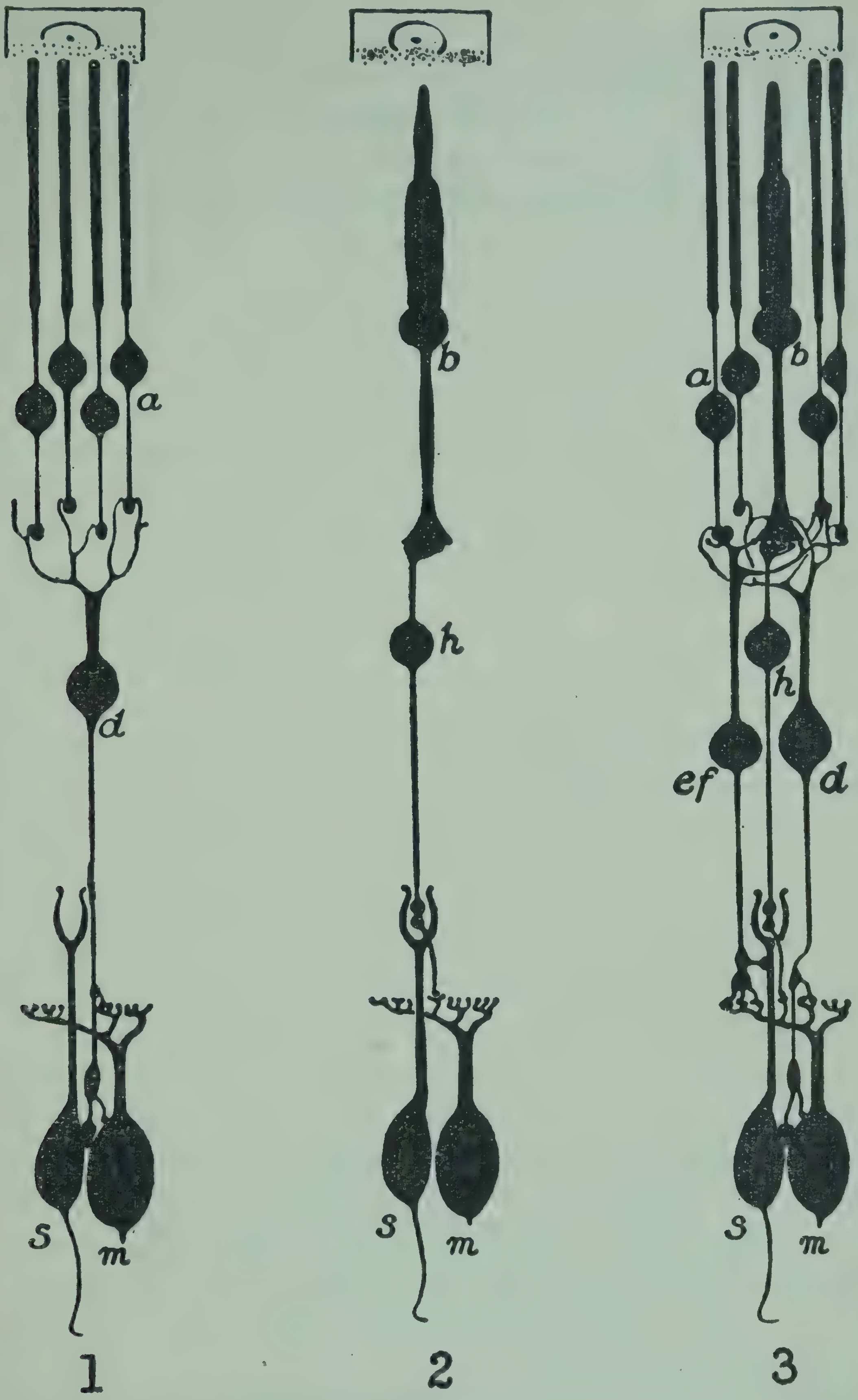


FIG. 39, 13

Grouping of the neurones of the primate retina into functional systems: (1) rod system with only one intermediate constituent, the mop bipolar (*d*); (2) pure cone system; (3) mixed or common rod and cone system. *a*, rods. *b*, cones. *d*, *ef*, *h*, centripetal bipolar cells. *s*, ganglion cells. *m*, parasol ganglion cells (S. Polyak (1941). *The Retina,* Plate 97. University of Chicago Press).

The examiner looks along the beam of light into the subject's eye and if his and the subject's eye are accommodated for infinity the optic disc and retinal vessels are clearly seen but the retina itself, being transparent, is invisible. The pigmented layer of the retina intervenes in front of the choroidal vessels so that the background of the fundus is a fairly uniform orange. The yellow colour of the macula cannot be seen but it can be distinguished by its position, two disc diameters on the temporal side of the disc, and by the absence of large blood vessels (Fig. 39, 9). At the optic disc where the nerve fibres leave the eye they become myelinated and therefore opaque.

The general structure of the retina is shown diagrammatically in Fig. 39, 12, but the detailed structure differs from place to place. At the fovea, cones only occur and each is connected, via first order neurones, to a single optic nerve fibre (Fig. 39, 13, 2). Fine detail can be appreciated at the fovea as the cones are extremely closely packed and the inner layers of the retina seem to be pressed away from the centre so that the light gains almost direct access to the foveal receptive elements. In the periphery of the retina light must first pass through the transparent fibre and ganglion layers before reaching the light-sensitive elements. The cone density rapidly falls from 125,000 per sq.mm at the centre of the fovea to about 6000 per sq.mm some 3° from the fovea and remains at this value up to the ora serrata. Rods begin to appear among the cones about ½° from the fovea and reach a maximum density of 150,000 per sq.mm. The central fovea is rod-free but no part of the periphery of the retina is cone-free. In the peripheral parts several rods are linked with one optic fibre (Fig. 39, 13, 1) so that discrimination of detail is not in fact very great. Since there are about 7×10^6 cones and 120×10^6 rods in the human eye but only 800,000 fibres in the optic nerve, a large number of receptors, on the average, must be connected to each optic nerve fibre. No rods or cones are present at

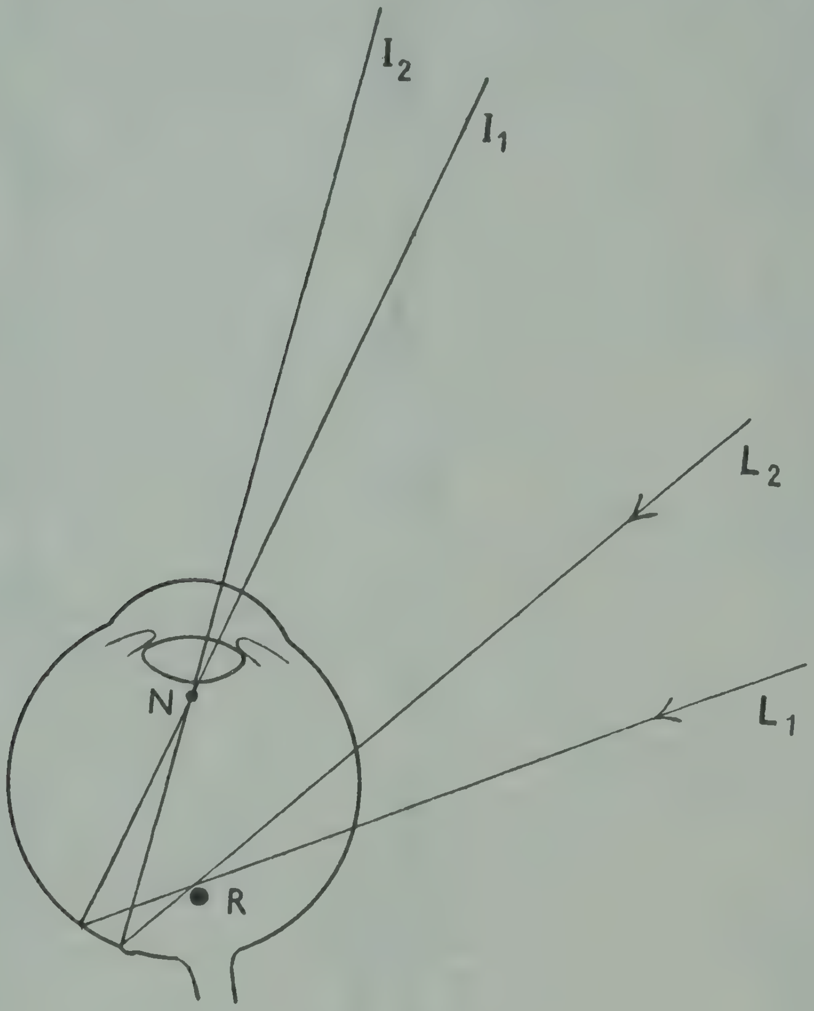


FIG. 39, 14

Diagram to illustrate Purkinje's experiment. A bright light is shone into the eye along L_1. The subject sees the edge of the shadow of the retinal vessel R at I_1. If the light is moved to L_2, then the image moves to I_2. N is the nodal point of the eye.

the optic disc so that this portion of the eye is blind. The external pigment layer of the retina is a single layer of hexagonal cells, with thread-like processes which pass between the limbs of the rods and presumably prevent the lateral spread of light in the retina.

The retina is obviously the light-sensitive layer of the eye in view of the effects of blockage of the retinal artery and on morphological grounds. It can be easily proved that the rods and cones are the receptive elements. On the optic disc and in front of the ora serrata there are no rods or cones and these areas are blind. Further proof is given by Purkinje's experiment which is best carried out in a dark room. A bright light projected through the sclera casts shadows of the retinal vessels on the retina. The observer looking at a distant wall sees dark red-brown branching shadows on a reddish background. If the source of light is moved by a given amount the images also move by an amount which can be measured (Fig. 39, 14). Since the optical constants of the eye are known, the distance between the retinal vessels and the receptive elements can be calculated. It is usually about 0·2 or 0·3 mm, which is the distance between the retinal vessels and the rods and cones as measured on a histological section.

VISUAL ACUITY

It has already been mentioned that, as tested with Snellen's types, normal visual acuity is sufficient to distinguish points which subtend an angle of one minute at the eye. This angle corresponds to about 4·5 μ on the retina and since the foveal cones measure about 3 μ in diameter it seems reasonable to conclude that two adjacent points of light can be distinguished if their images fall on two separate cones with an intermediate unstimulated cone. But in special circumstances, such as in a Vernier scale where two lines have to be brought into coincidence, the acuity is much higher and a difference of less than ten seconds of arc can be recognized. In this case the images of the lines are laid across the pattern or mosaic of cones, and although the image of a line may fall on two adjacent cones one is likely to be stimulated more than the other and will send a greater number of impulses to the brain. The larger number of such clues given by a line presumably allows greater differentiation than do two points of light. In a very dim light (scotopic vision) visual acuity is poor probably because groups of rods are connected to a single fibre of the optic nerve (see p. 699).

In dealing with practical problems of illumination it is important to choose appropriate light intensities. Illumination may be quite low in corridors but, since acuity improves with increased brightness, rooms in which fine work is to be carried out require good lighting.

DUPLEX THEORY OF VISION

Two quite different mechanisms seem to be involved in vision. One of these, called photopic, operates in bright light and is concerned with colour vision; the other, called scotopic, is concerned with vision in a dim light and does not call forth colour sensations. Photopic vision is mediated by the cones and scotopic vision by the rods. The evidence for this is given below and in the section on dark adaptation (p. 696).

If a bright spectrum produced by means of a prism is examined it will be seen to be coloured as indicated in Fig. 39, 15, which shows also the wavelengths of the visible rays. The maximum brightness is in the orange region at 0·61 μ. If the intensity of the spectrum is greatly reduced all colour is lost and it appears as a luminous band with its brightest region about 0·535 μ (seen as green in a bright spectrum). At the same time the red end of the spectrum becomes very dark. (A prism does not produce an equal energy spectrum and so the wavelengths of the brightest regions differ from those given in the next paragraph.) Photopic vision is best at the fovea centralis where cones only are present; if an object is seen by the fovea it is seen in colour. When the light intensity is reduced the fovea is blind, but the more peripheral parts having a lower threshold are still able to function.

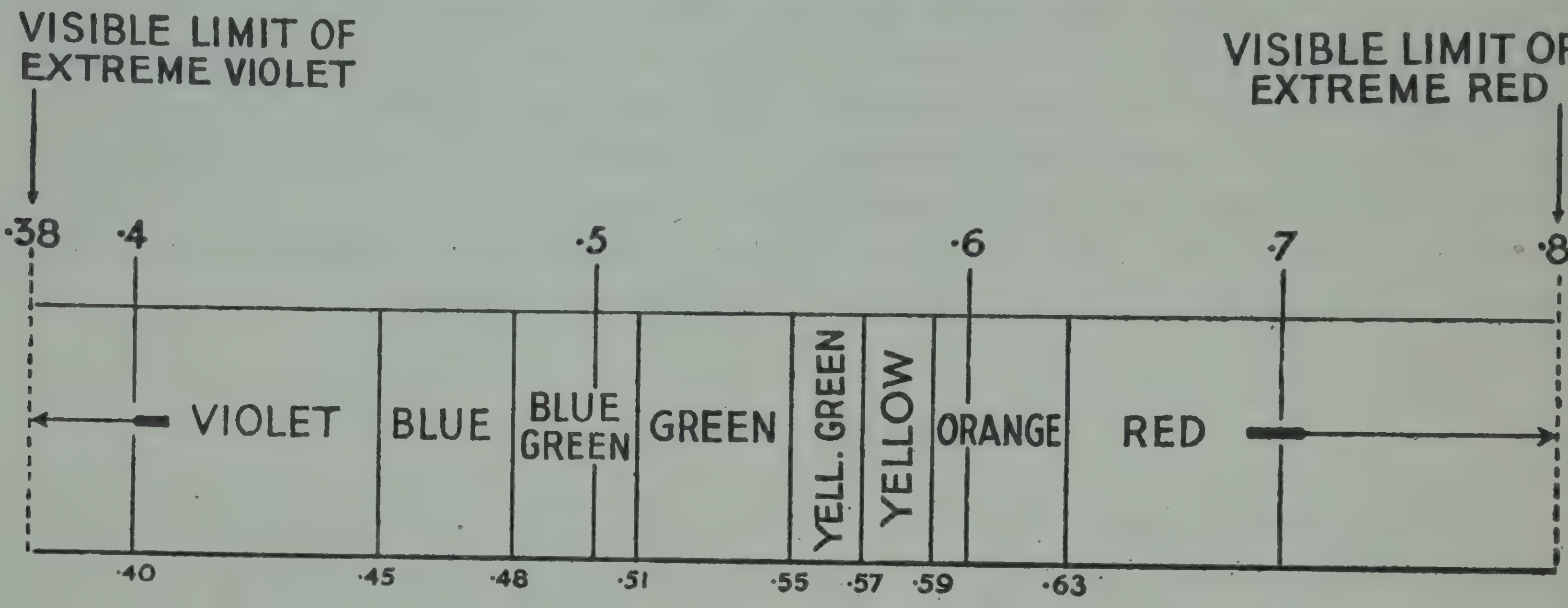


FIG. 39, 15

Diagram to show the colour names given to the various parts of the spectrum. Wavelengths are given in μ. Compare with the spectra shown in Fig. 25, 1.

By measuring the energy required at each wavelength to produce a response of constant magnitude it can be shown that the spectral sensitivity of the two mechanisms is quite different. Obviously if a large amount of energy is required the receptor is insensitive, and if only a small amount is required the receptor could be described as sensitive. In other words, sensitivity is indicated by the reciprocal of the energy needed to produce the standard response; when this is plotted against wavelength we get an 'equal energy luminosity curve.' If the measurement is carried out at high light intensity (cone vision) the maximum sensitivity falls at 0·555 μ: for low intensity it is at 0·51 μ (rod vision). The maxima in both cases have been taken as 1 in Fig. 39, 16. Because the scotopic (rod) mechanism is extremely insensitive to red, red glasses were worn during the day by pilots of night fighters to prevent light adaptation and loss of sensitivity to dim lights. The rod luminosity curve has a remarkable resemblance to the absorption curve of visual purple, which has its maximum at 0·502 μ. Visual purple is found only in association with the rods. There is some overlap in function between the rod and cone mechanisms since the photopic (cone) sensitivity curve given in Fig. 39, 16, applies to brightness down to 1 millilambert, whereas the scotopic (rod) curve applies to brightness below 10^{-3} millilamberts. Intermediate curves are obtained between 10^1 and 10^{-3} millilamberts.

The visible spectrum (Fig. 39, 15) is limited to radiations of wavelength 0·4 to 0·7 μ. As there is little absorption of infra-red light by the transparent media up to 1·1 μ the red limit must be due to the failure of the retina to respond to light in the range between 0·8 and 1·1 μ. The other limit at the violet end is due to the marked absorption of ultraviolet light by the transparent media and not to lack of sensitivity of the retina since removal of the crystalline lens renders the ultraviolet rays more visible. Ultraviolet radiation, because it makes the proteins of the lens fluoresce, produces a diffuse illumination within the eye. X-rays may also be detected by the eye possibly by causing the retina to fluoresce. This property of the retina has been used to measure the axial length of the eye (p. 680).

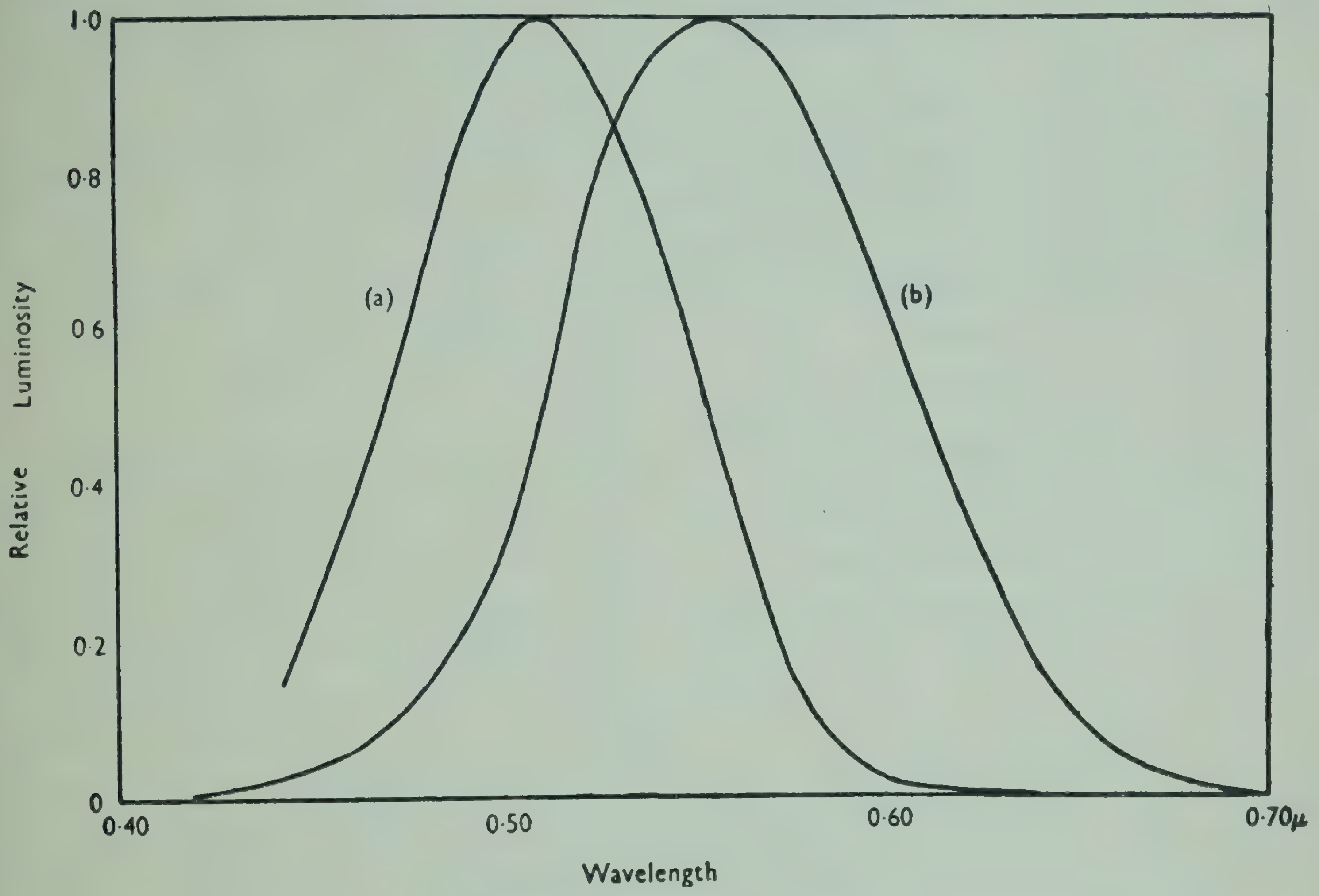


FIG. 39, 16

The equal-energy luminosity curve (*a*) for rod vision, as measured at low intensities in the extra-foveal area of the retina; (*b*) for cone vision, as measured at high intensities at the fovea (W. D. Wright (1944), *The Measurement of Colour*. London: Hilger).

FIELD OF VISION

The extent of the field of vision, that is the area projected on to the inner surface of a sphere, in which a luminous object can be seen, is measured by means of a *perimeter*. The subject covers one eye and looks with the other at a fixation point at the centre of the apparatus while a white disc is brought in from the periphery of the field of vision until the subject indicates that he is able to see the disc. The shape of the field (Fig. 39, 18) is dependent to some extent on the shape of the face, since it is restricted on the medial side by the nose and above by the supra-orbital margin. Laterally there is no obstruction to vision and with the eye looking straight forward in the head objects up to about 100° away from the visual axis can be seen. If the eyes are moved right and left the total field of vision around the head becomes quite considerable so that the

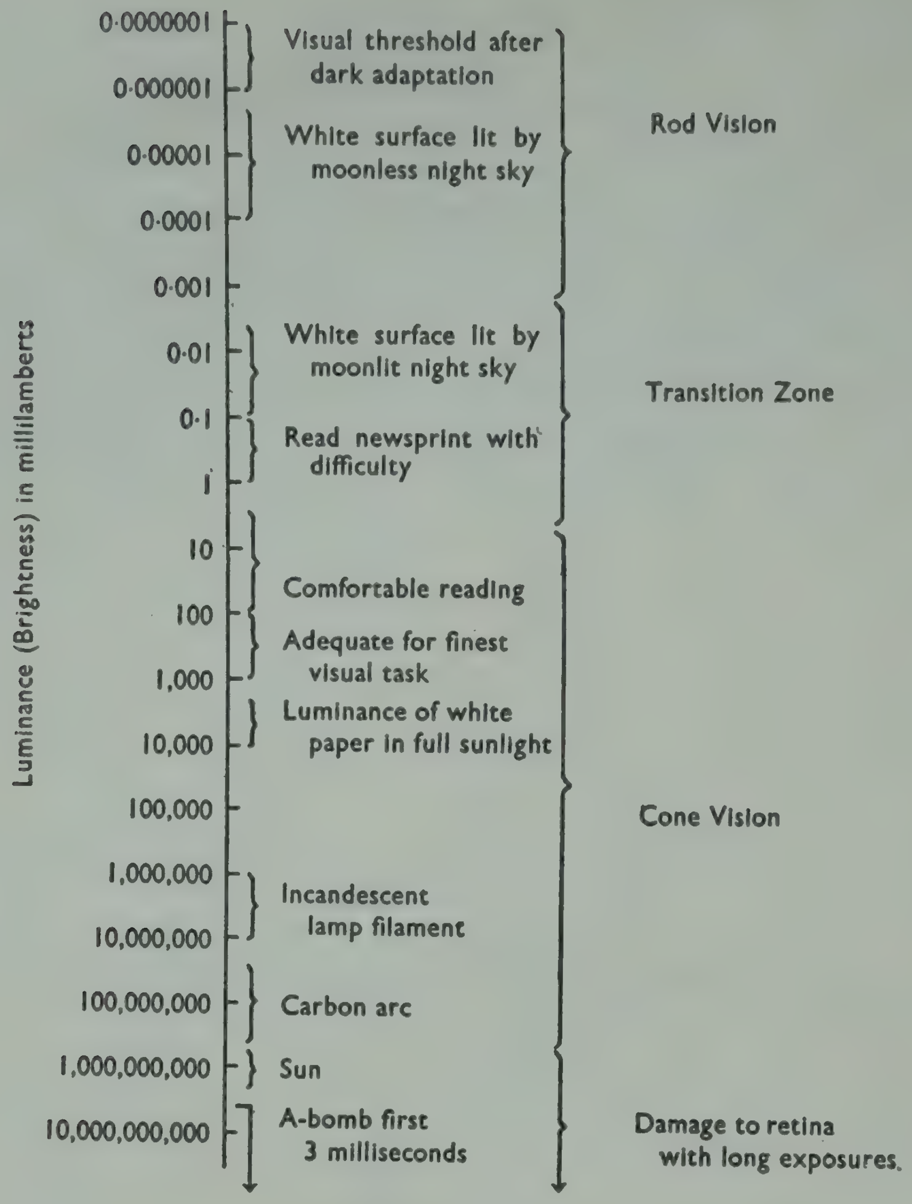

Luminance (Brightness) in millilamberts
0·0000001
0·000001
0·00001
0·0001
0·001
0·01
0·1
1
10
100
1,000
10,000
100,000
1,000,000
10,000,000
100,000,000
1,000,000,000
10,000,000,000
Visual threshold after dark adaptation
White surface lit by moonless night sky
White surface lit by moonlit night sky
Read newsprint with difficulty
Comfortable reading
Adequate for finest visual task
Luminance of white paper in full sunlight
Incandescent lamp filament
Carbon arc
Sun
A-bomb first 3 milliseconds
Rod Vision
Transition Zone
Cone Vision
Damage to retina with long exposures.

FIG. 39, 17

Range of luminance (brightness) to which the eye may be subjected with indications of the receptive mechanisms involved.

(*By courtesy of F. W. Campbell.*)

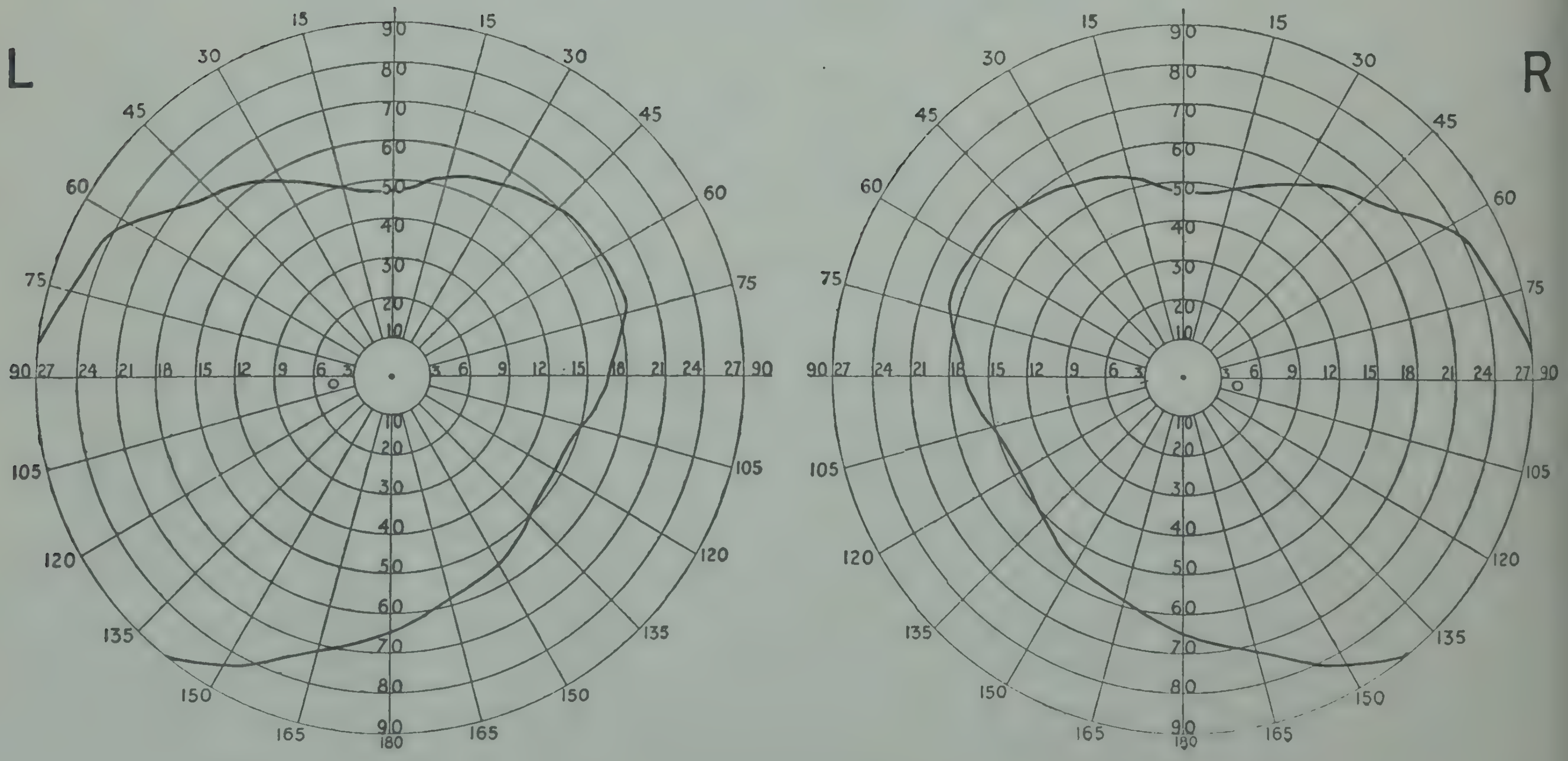


FIG. 39, 18

The field of vision.

blind zone behind the head may be as little as 140°. Such a large field, however, is obtained only under somewhat artificial conditions. Objects far from the centre of the field cannot be easily recognized, since the acuity there is small and since colours may not be perceived. If coloured discs are used to plot out the field of vision the subject can see the discs in the extreme periphery of the field without being able to recognize their colours until the discs are brought nearer the centre. The fields for colours are roughly concentric with the field for white; red and green usually have the smallest field and blue the largest. Beyond this point shades of grey only are seen. But there is no hard-and-fast boundary to these different regions since the field of vision for any one colour depends on the brightness of the coloured disc used to map it out. If the light reaching the periphery is sufficiently intense its colour can be recognized. The difficulty in appreciating colour in the periphery is due apparently to the small number of cones present per unit area as compared with the very large numbers occurring in the fovea.

Colour Vision

The Phenomena of Colour Vision.—The eye differs from the ear in that the information which it provides about its stimulus concerns not only intensity and quality, and the way in which these change with time, but also, and in great detail, the way in which they are distributed in space. It provides this immensely greater information on spatial distribution at the cost of giving very much less on quality.

The ear (see pp. 715–720) supplies the brain with a harmonic analysis of any sound that stimulates it, so that to specify the sensation produced even by a steady musical note, it may be necessary to fix the values of several thousands of independent quantities, the harmonic components. Two physically different sounds cause identical sensations only if the sounds are alike in the values of all these quantities. For the eye, on the other hand, two light stimuli, even if widely different in their composition as revealed by spectroscopic examination, can cause identical sensations provided that they are alike in the values of only *three* independent quantities, suitably defined. This threefold nature of colour vision, now generally known as trichromacy, was first explicitly pointed out in 1756 by the Russian scientist, M. V. Lomonosov, who wrongly attributed it to the nature of light itself. It was first appreciated as a property of the eye by Thomas Young (see p. 693).

The trichromacy of colour vision may be illustrated by the familiar fact that the stage lighting engineer, provided with red, green and blue lights, can, by mixing them in suitable proportions, exactly reproduce almost any other colour. In fact, although any pale or *unsaturated* colour can be exactly reproduced, the match is only approximate for some of the pure spectral or *saturated* colours (‘saturation’ being the technical name for freedom of a colour from admixed white). An exact match can, however, always be obtained if one of the three primary colours may be added not to the other two, but to the test colour which it is required to match. Thus, for example, pure spectral blue-green cannot be exactly matched with a mixture of spectral blue, green and red, or of spectral blue and green alone, but a mixture of spectral blue-green and red can be exactly matched with a mixture of spectral blue and green. Spectral yellow can be very nearly matched with a mixture of red and green, but to make the match perfect, a little blue must be added to the spectral yellow. If the addition of a light to one of the two fields on which the matching is being carried out is conventionally represented as equivalent to its subtraction from the other, then the amounts, positive or negative, of three specified primaries required to match a given light provide a set of three quantities which define uniquely the sensation which it produces.

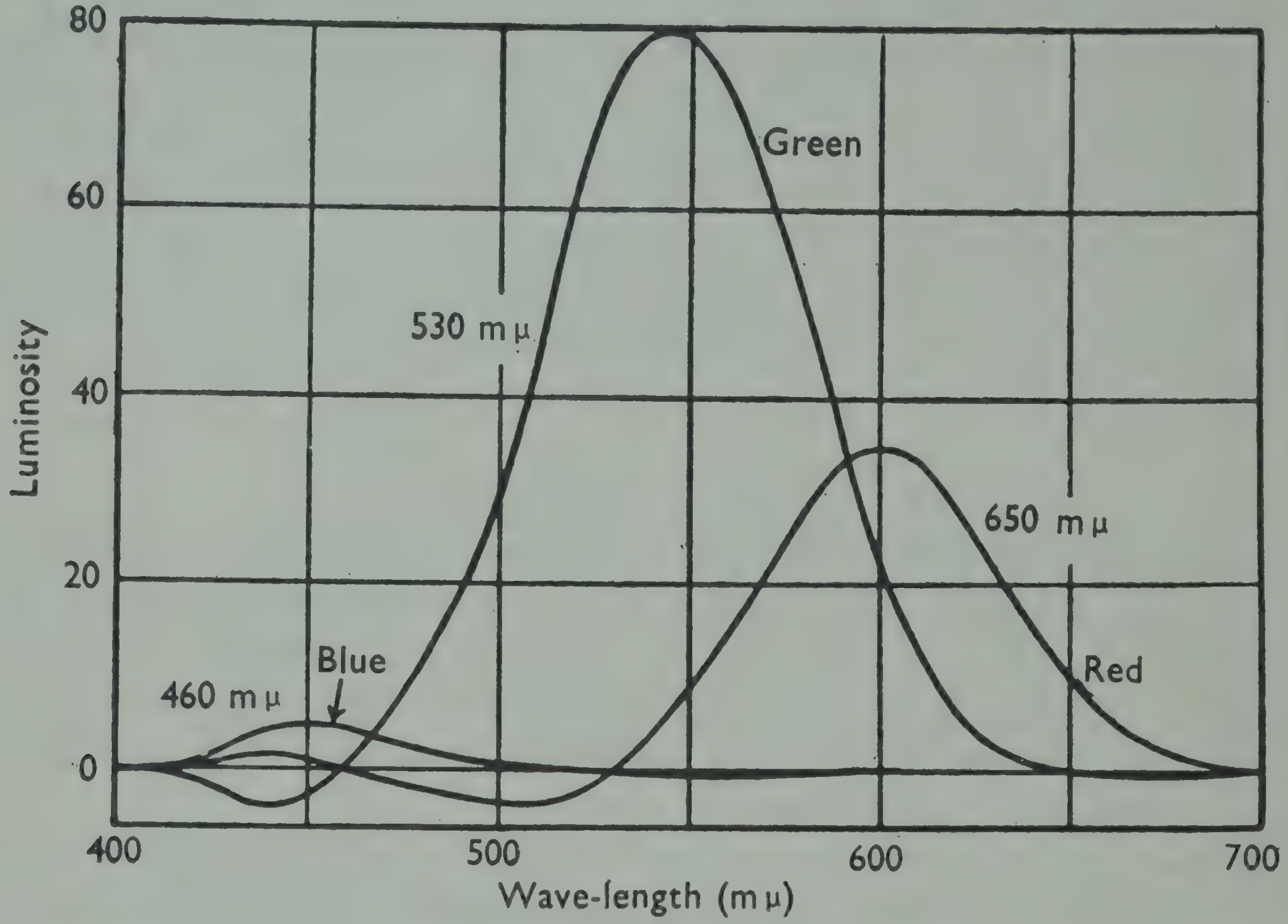


FIG. 39, 19

Colour mixture data for the standard spectrum, incorporating the photopic luminosity data. (*After* W. D. Wright (1928), *Trans. opt. Soc.* **30,** 141. From A. F. Rawdon-Smith (1938), *Theories of Sensation*. Cambridge University Press.)

Fig. 39, 19 shows the result of measuring these quantities for monochromatic lights of all spectral wavelengths, using red, green and blue primaries. The primaries used need not necessarily be red, green and blue, but these have the advantage as primaries that the quantities of them required to match other colours are mostly positive, and when negative, are small.

The phenomena of colour mixture were described by Grassmann in 1853 by the following laws:

1. Every colour sensation can be specified by three mathematically definable quantities. (In Grassmann's formulation these were the hue, the intensity and the saturation. They could equally well be the amounts of three specified primaries in a colour match).
2. If one of two components in a colour match is altered continuously, the other remaining constant, then the appearance of the mixture alters continuously.
3. If two colours are alike in hue, intensity and saturation, then they behave alike in colour mixture, however much they may differ in spectral composition.

4. The intensity of a mixture is the sum of the intensities of its constituents.

Modern research has shown that Grassmann's first law is exactly true provided that the area of retina illuminated is of moderate size, corresponding to an external field of between about 20′ and 3° diameter. There are departures from it for very small (see p. 693) and very large fields. The second law is exactly true. The third is true to a close approximation, though very small departures from it have been found. The validity of the fourth law depends on the method of measuring intensity. The method which makes it most nearly true is flicker photometry, where two lights are considered to be of equal intensity when they have been adjusted to give the least possible flicker when alternated one with the other.

Complementary Colours.—A consequence of Grassmann's laws of colour mixture which can easily be verified experimentally is that certain pairs of spectral colours when mixed give white. These are called complementary colours. Thus, yellow is complementary to blue, and red to blue-green. Not all the colours of the spectrum have complementary colours also in the spectrum. The complementary colour to green is magenta or purple, which is not to be found in the spectrum but may be made by mixing spectral red and spectral violet.

Successive and Simultaneous Contrast.—If a bright source of white light is looked at for some time and the gaze is then transferred to a black area, a faint white image is seen of the same apparent size and shape as the original source. This is called a *positive after-image.* If the gaze is transferred to a uniform white area of suitable brightness, a dark or *negative* after-image is seen. If the original source is coloured, its positive after-image appears of approximately the same colour, and its negative after-image of approximately the complementary colour. After-images are often referred to as *successive contrast phenomena.*

A small neutral grey patch on a white background appears darker than a similar one on a black background. If placed on a strongly coloured background, such a small grey patch appears tinged with the complementary colour. These changes in the appearance of a field according to the brightness or colour of its surroundings are known as *simultaneous contrast phenomena.*

Retinal Directional Effect.—If light falls on a given area of the retina from the periphery of the pupil, it appears less bright than light of the same intensity falling on the same spot on the retina from the centre of the pupil. The sensitivity of the retinal receptors thus depends on the direction from which light reaches them. This phenomenon is called the retinal directional effect or Stiles-Crawford effect. It is large for the cones of the fovea, but very small for the dark-adapted peripheral retina tested with dim light, that is, when rods only are active.

Colour Blindness.—Abnormalities of colour vision of genetic origin, generally known as colour blindness, are common in men, appearing in about 8 per cent. of the male population, but much less common in women (0·4 per cent.). The common types are inherited as totally sex-limited recessive characters, which are probably all allelomorphic.

Colour-blind subjects fall into a number of fairly sharply distinguishable categories. *Monochromats* are quite unable to distinguish colours: any colour can be matched by them with any other, if the intensities are suitably adjusted. Monochromats can be divided into *rod monochromats*, who have a luminosity curve like that of the normal dark-adapted eye (see Fig. 39, 16), and *cone monochromats*, whose luminosity curve is closer to that of the normal light-adapted eye. Both types are rare, the cone type especially so. *Dichromats*, who can match all colours with suitable mixtures of two primaries, are divided into *protanopes* and *deuteranopes*, forming about 1·0 per cent. and 1·2 per cent. of the male population respectively, and *tritanopes*, who are much rarer, and nearly as often female as male. The protanopes and deuteranopes are often grouped together with some of the anomalous trichromats as 'red-green blind.' They have very little colour discrimination at the red end of the spectrum, and hence confuse red, brown and green objects, though yellow ones are easily distinguished owing to their higher reflecting power. Protanopes and deuteranopes differ from each other in the form of their photopic luminosity curves. Tritanopes, sometimes misleadingly called 'blue-blind,' have normal colour discrimination at the red end of the spectrum, but very little ability to distinguish blue from green. *Anomalous trichromats*, forming nearly 6 per cent. of the male population, resemble normal subjects in that they require three primaries to match all colours by colour mixture but they require them in abnormal proportions so that they neither accept normal colour matches nor make colour matches which normal subjects accept. They have usually also a defect of colour discrimination at the red end of the spectrum, but not so severe a defect as have the protanopes and deuteranopes. The theoretical basis for these findings is discussed further on the opposite page.

Colour-blind subjects, even when their colour discrimination is severely defective, are often unaware of their defect. In familiar situations they compensate for their defective colour discrimination by an increased use of alternative clues, based in part on a prior knowledge of the usual colours of objects which they recognize from their shape. It may be only on the rare occasions when these alternative clues are entirely absent that the defect becomes obvious.

The detection of colour blindness is important in selecting engine drivers and others who must distinguish coloured signal lights. A convenient quick test consists of 'pseudo-isochromatic plates,' of which Ishihara's are the most widely used. Each plate is a card on which an irregular array of coloured dots is printed, so that a letter or figure is formed by the dots of one colour, other colours forming the background. Some cards are designed to be read easily by normal subjects but not by the colour-blind, others to be read more easily by the colour-blind than the normal, others again to be interpreted differently by the normal and the colour-blind. This test is very efficient at separating the normal from the abnormal, but does not distinguish very accurately between the different types of abnormality. To decide whether a subject with a mild abnormality can safely be employed in any given occupation requiring discrimination of colour, a special test, imitating the task which he will have to carry out, is often used.

Though hereditary colour-blindness is vastly commoner, defects of colour vision can also be acquired as a result of disease of the retina, optic nerve or visual cortex, or of tobacco poisoning. These acquired defects are usually accompanied by severe defects of visual acuity or of visual fields.

Colour Discrimination with Small Central Foveal Fields.—If a pair of colorimetric fields subtending only 20′ or less at the eye is carefully fixated with the central fovea, it is found that all colours can be matched with mixtures of only two primaries. The normal (trichromatic) subject is thus dichromatic for small central foveal fields, the type of dichromacy being very similar to tritanopia. Protanopes and deuteranopes, tested in the same way with small central foveal fields, are found to be monochromatic. The special features of colour discrimination with small central foveal fields are in part a local property of the central fovea and in part a result of the smallness of the field: dichromacy can also be obtained with small fields which are not centred on the central fovea, but the size of field required is slightly smaller, and matching is less precise.

Theories of Colour Vision.—Thomas Young in 1801 suggested that the trichromacy of normal vision could be explained if there were three kinds of receptors active in the retina; and that protanopia, the only kind of colour blindness then known, might be due to the absence of those receptors most sensitive to red. During the past 150 years a great number of theories of colour vision have been published, most of them accepting the principle of Young's suggestion, but modifying or supplementing it, a few abandoning it altogether and leaving the basic fact of trichromacy wholly unexplained. Most of these theories are now only of historic interest.

A typical modern working hypothesis, incorporating Young's suggestion and also the duplicity theory (see p. 685) is as follows. Scotopic vision depends on rods only and, since these are all of one kind, scotopic vision is monochromatic. Human photopic vision depends on cones only, and these are of three kinds, each kind having its own special photosensitive pigment and hence its own characteristic spectral sensitivity curve. The so-called 'red' receptors have their maximum sensitivity somewhere near $0{\cdot}57\ \mu$ (in the yellow or possibly the orange part of the spectrum), the 'green' receptors near $0{\cdot}54\ \mu$, and the 'blue' receptors near $0{\cdot}45\ \mu$. Protanopia is due to absence or inactivity of the 'red' receptors, or to their containing the pigment appropriate to one of the other kinds, deuteranopia to similar absence or abnormality of the 'green' receptors, and tritanopia of the 'blue.' Anomalous trichromacy can have no such simple explanation; presumably at least one of the receptor pigments of the anomalous trichromat is chemically different from its normal counterpart.

The above hypothesis is simple, is not contradicted by any of the well-established facts of colour vision, and provides, either alone or with simple and likely supplementary assumptions, direct explanations of most of the facts. There are, however, other hypotheses which, though indirect evidence may make them appear less likely than the one first stated, cannot be regarded as disproved. It has, for example, been suggested that the three kinds of receptors required to explain trichromacy are not three kinds of cones, but two kinds of cones and one of rods. If so, these rods cannot, at least in the foveal region, use visual purple as their photosensitive pigment, since colour matches for a 2° central field are often far from being matches for the visual-purple-using rods of the periphery, even when allowance is made for the screening effect of macular pigment. Two other hypotheses perfectly consistent with the facts, though lacking definite supporting evidence, are that there are not three kinds both of cones and of pigment, but *either* three kinds of cones, and more than three of pigment, *or* three pigments and more than three kinds of cones. These hypotheses require that some at least of the cones contain mixtures of photosensitive pigments. Yet another possibility is that there are more than three kinds of pigments and of cones, and that vision is trichromatic because at some more central stage in the visual pathway, possibly the retinal ganglion cells or the lateral geniculate body, there are cells of only three kinds in respect of spectral sensitivity. Against this last hypothesis there is strong though not absolutely compelling indirect evidence: it is very difficult to reconcile with Grassmann's third law, with the stability of colour matches to moderate degrees of light adaptation, and with the manner in which they are disturbed by adaptation to very bright lights.

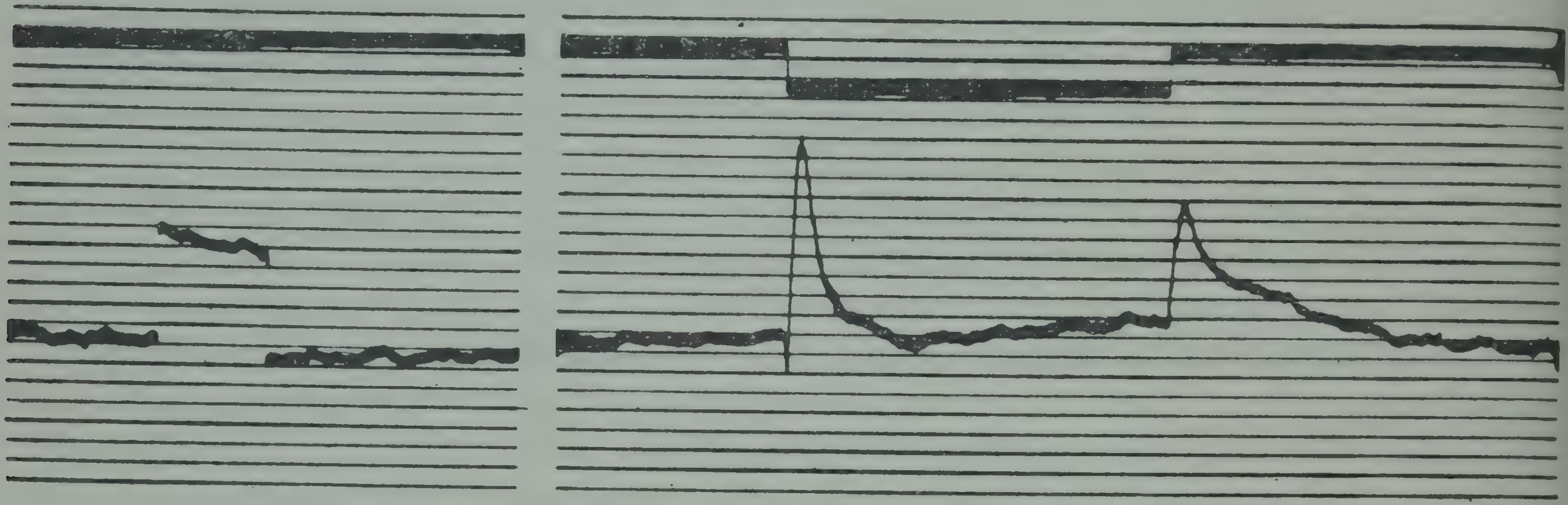


FIG. 39, 20

Electroretinogram of a mammalian eye (squirrel). Upward deflexion represents the front of the eye becoming positive to the back. Calibration on the left 500 μV. Duration of stimulus 1 sec. The initial downward deflexion is called the *a*-wave, the first upward deflexion the *b*-wave, and the second upward deflexion, occurring after the stimulus is extinguished, the *d*-wave (off-effect). (Arden, G. B. & Tansley, K. (1955). *J. Physiol.* **127,** 595.)

ELECTRICAL ACTIVITY OF THE RETINA

Two kinds of electrical activity in response to changes in illumination can be recorded from the retina. If one electrode is placed on the cornea and the other on any other part of the body, a change in potential of the form shown in Fig. 39, 20 is found in response to illumination. This is called the *electroretinogram.* In some invertebrate eyes, the electroretinogram is known to be produced solely by the photoreceptors, and there is some evidence that this may be true also of the vertebrate eye.

Responses of another layer of the retina, the layer of ganglion cells, can be recorded with a micro-electrode placed directly on the exposed inner surface of the retina. Such an electrode, if sufficiently small and correctly placed, picks out the spike potentials produced by a single ganglion cell.

Some ganglion cells, called on-elements, discharge impulses during illumination but not in darkness, some (on-off-elements) for brief periods at the beginning and after the end of illumination, and some (off-elements) only

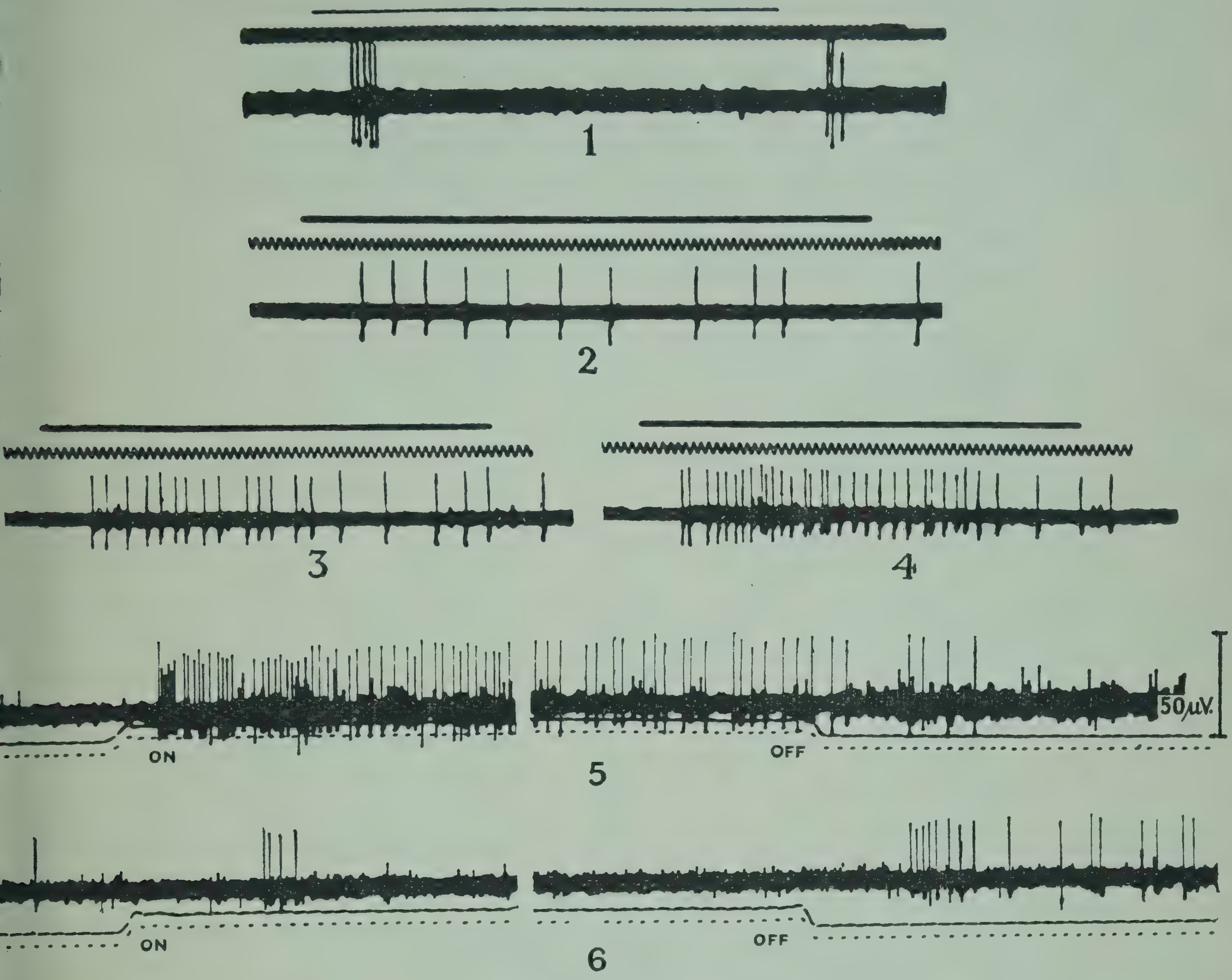


Fig. 39, 21

1 to 4. Impulses picked up by a micro-electrode placed on the inner surface of the retina. Above each oscillogram is the time record (50 per sec) and light signal. 1. 'On-off' element from photopic cat's retina responding to light of wavelength 0·660 μ. Somewhat above threshold. 2–4. 'On'-element from scotopic retina of guinea-pig responding to 'white' light: 2, at strength 0·006 m.c.; 3, at 0·018 m.c.; and 4, at 0·061 m.c. Note increasing frequency and shortening latent period, as stimulus intensity increases (R. Granit (1943), *Nature*, **151,** 10). 5 and 6. Single fibre discharge from the optic nerve of a rabbit obtained by the use of a silver-filled microcapillary. In 5, the fibre shows a brief discharge after a latent period of 0·04 sec when the light is switched 'on' and then a continuous discharge which gradually declines in frequency until the light is turned 'off.' There is then a short pause which is followed by the 'off' discharge of four spikes. In 6, another fibre shows a discharge at 'on' (the first spike in this record is spontaneous) and a number of spikes at 'off.' There is a break of 0·4 sec in the middle of records 5 and 6. The time dots are 10 msec apart (L. C. Thomson (1953), *The Medical Press*, **230,** 3).

after the end of illumination. These types of response are shown in Fig. 39, 21. The complexity of the pattern of response does not necessarily imply that the responses of the rods and cones are equally complex; it may be due in part to interaction in the network of retinal neurones lying between the rods and cones and the ganglion cells.

Mechanism of Dark Adaptation

The eyes take quite a long time to become accustomed to very low illumination after having been exposed to a high one. For example, a luminance of 0·01 millilamberts is just visible after switching off a very bright light; an hour or so later in the dark the smallest visible intensity may have fallen to 0·000001 millilamberts (a decrease of 10,000 times). This increase in sensitivity or dark adaptation goes on for at least an hour.

The curve of Fig. 39, 22, was obtained by exposing the observer to a very bright light for some minutes. The light was then turned out and the visual threshold, that is the smallest detectable luminance, was measured from time to time during the following sixty minutes. The curve has a kink in it which suggests that there are two mechanisms in the retina adapting at different rates. The first part, lasting 6 to 11 minutes, is characteristic of the photopic mechanism which is presumably mediated by cones; the later scotopic part of the curve, mediated by rods and visual purple, seems to have an initial delay in adaptation. In night-blind persons, in whom the scotopic mechanism is absent, the threshold declines very slowly to about 500 micromillilamberts as shown by the dotted line in Fig. 39, 22.

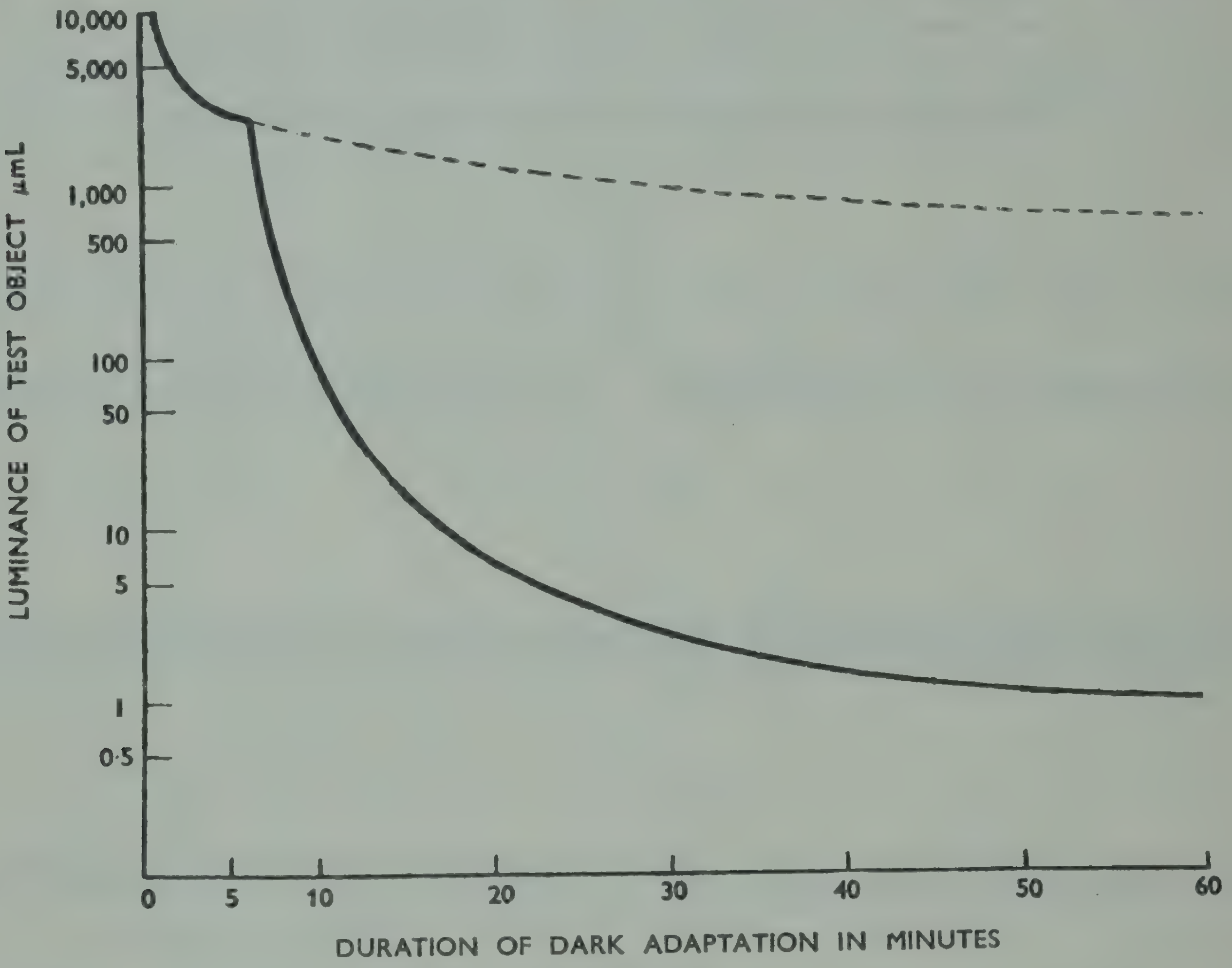


Fig. 39, 22

Horizontally: duration of dark adaptation in minutes. Vertically: the smallest visible luminance in micromillilamberts. The continuous line shows the course of dark adaptation in normal persons; the dotted line refers to night-blind persons. (*After* Kohlrausch.)

The shape of the dark adaptation curve is dependent on the angle subtended by the test object and on the part of the retina which it stimulates. A small field acting on the foveal region shows no kink in the curve; extrafoveal regions always show the kink. In the dark adapted eye a small field in the extrafoveal region shows a higher threshold than a large field which stimulates large groups of receptors. This difference is not easily explained on the basis of changes in visual purple but can be interpreted as evidence of retinal integration as envisaged by Lythgoe (p. 699). This view is supported by the finding that the foveal and extrafoveal thresholds of the dark adapted eye are very nearly the same if the test object subtends an angle (say 2·7′) approximately that subtended by a rod or a cone. This finding appears, however, to conflict with a well-known astronomer's trick. Faint stars in the night sky, too faint to be seen if directly looked at, may be seen if the gaze is directed a few degrees away. But in this 'indirect fixation,' eye movements occur which move the image of the star across various groups of receptors and by the nervous integration of the retina of the dark adapted eye a subthreshold stimulus may rise above the threshold and be seen.

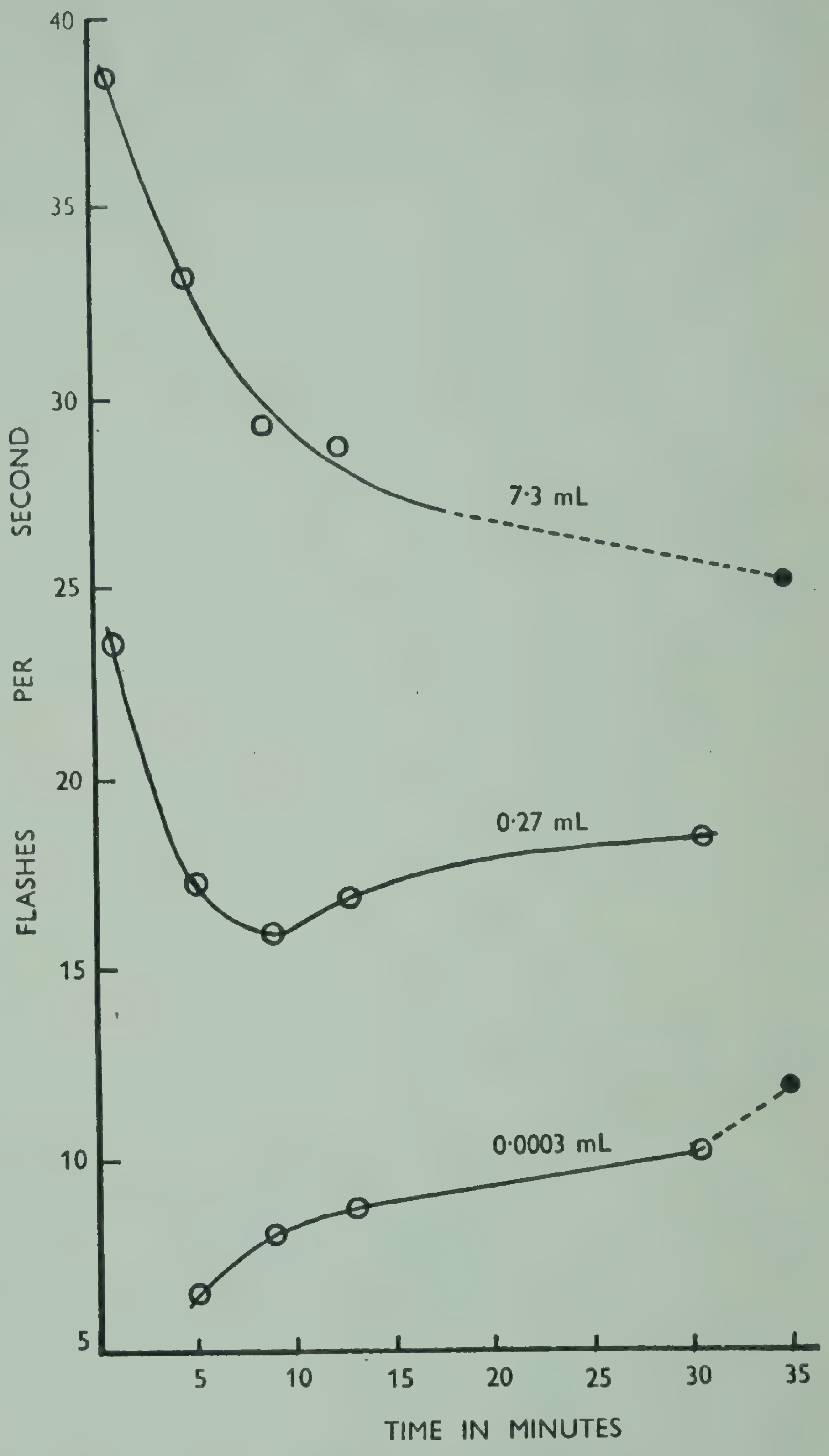


Fig. 39, 23

To show the changes in the critical frequency of flicker during dark adaptation. For any one curve the luminance of the test patch was unaltered (see values set against each curve). All observations were made with 10° of eccentric vision. ● After 60 minutes in the dark. (*After* R. J. Lythgoe (1940), The mechanism of dark adaptation. *Brit. J. Ophthalmol.* **24,** 25.)

If a constant source of light is interrupted by a slowly rotating shutter a flickering sensation is produced, but if the speed of rotation is increased sufficiently the flickering sensation disappears and the light appears to be constant. The sensation of light lasts a little longer than the actual stimulus and so fills in the gap between successive flashes of light. The minimum rate of interruption which just removes the flickering sensation is called the critical frequency of flicker. Its variations during dark adaptation can be obtained in the following way. The subject sits in a brightly lit cube looking at a flickering test patch of light in an opening in one wall; the patch is shown and obscured for equal lengths of time by a rotating shutter. The critical frequency is found and then the light in the cube is switched off and further measurements are made during dark adaptation, the brightness of the test patch being kept constant through any one experiment (Fig. 39, 23). For illuminations stimulating the scotopic mechanism only (0·0003 mL for example) a rise of critical frequency is observed in dark adaptation; with greater brightness of the test patch (such as 7·3 mL) which stimulates the photopic mechanism a fall of critical frequency occurs; at intermediate values there is a fall followed by a rise. Thus, dark adaptation is characterized not only by fall in visual threshold but by other changes which modify both scotopic and photopic vision and give further support to the duplex theory of vision.

There is no definite brightness at which there is a sharp change-over from photopic to scotopic vision for a gradual transition occurs from predominantly cone vision to predominantly rod vision (Fig. 39, 17). This transitional zone is about 0·01 mL which is the brightness of a piece of white paper viewed by the light of the full moon.

The fully dark adapted retina is a remarkably sensitive detector of light energy. Indeed, to express its sensitivity in physical units correctly involves the use of the smallest subdivision of energy—the quantum. A quantum is the smallest discrete 'parcel of energy' which can be transferred from one system to another. The work of Pirenne has shown that a light sensation arises if one quantum of light of wavelength 0·51 μ is absorbed every second for 5000 rods. There is strong evidence that if one quantum of light is absorbed by each of from 2 to 10 rods all within about 1/5 second the signals from these rods summate and initiate a nervous response which reaches consciousness.

The bleaching of visual purple in the light and its subsequent regeneration in the dark gives us a physical basis for the phenomena of dark adaptation. For example, it might be assumed that in order to produce a minimal stimulus a certain amount of visual purple has to be broken down; if this were true then the greater amount of visual purple present in the retina the smaller would be the illumination necessary to produce this amount of breakdown, or in mathematical terms for a threshold stimulus—

$$\text{Illumination} \times \text{Concentration of visual purple} = \text{constant.}$$

In the case of the cat the concentration of visual purple increases between the 10th and 30th minute of dark adaptation from 57 to 95 per cent. of the final value. If the concentration of visual purple increases in the human retina in a similar fashion the visual threshold should be reduced to about a half according to the above equation, but it is reduced in Fig. 39, 22 from 100 to 2·5 mL, that is about 40 times.

It is also difficult to attempt a quantitative explanation of vitamin A deficiency in which there is lowering of light threshold and retarded regeneration of the visual purple, since the amount of visual purple in the deficient eye is greater than would be expected from the equation given above. But in vitamin A deficiency the picture is complicated by degeneration of rods and ganglion cells.

Visual purple is a complex material formed by the combination of a protein with *retinene*, vitamin A aldehyde (Fig. 10, 5). Various breakdown products of visual purple are known from which the pigment can be regenerated. This process is affected by the previous illumination, being greater in animals previously exposed to bright light. This, and the fact that anoxæmia slows the re-formation of visual purple (p. 119), suggests that energy is required for regeneration. Even a minor degree of anoxia, such as is encountered in flying at 5000 ft., detrimentally affects night vision by contracting the field of vision.

The sort of change envisaged by Lythgoe to explain the great increase in sensitivity during adaptation is that each fibre of the optic nerve serves several elements by a spread of synaptic connexions (Fig. 39, 24). This would account both for the observed reduction in the fineness of discrimination and for the increased retinal sensitivity to light during dark adaptation. This theory of increasing synaptic connexions is by no means improbable, since it has been shown that nervous interactions can occur in the retina. The discharge obtained by switching on and off a spot of light directed on the receptive field of a ganglion cell (see p. 695) can be inhibited by switching on and off simultaneously a second spot of light which falls outside the receptive field of the first. Retinal interaction can also be invoked to explain the effect on the acuity of the fovea of a light stimulus falling on the peripheral retina. This well-known phenomenon is described thus in *The Merchant of Venice*, "when the moon shone, we did not see the candle . . . so doth the greater glory dim the less." A neighbouring source of glare depresses the apparent brightness of a light source in one-tenth second and keeps it so for the duration of the experiment. This process cannot be a chemical one because it takes place so quickly, and because it takes place at a distance. A glaring source of light shortens the latent period of vision for another light exhibited near by.

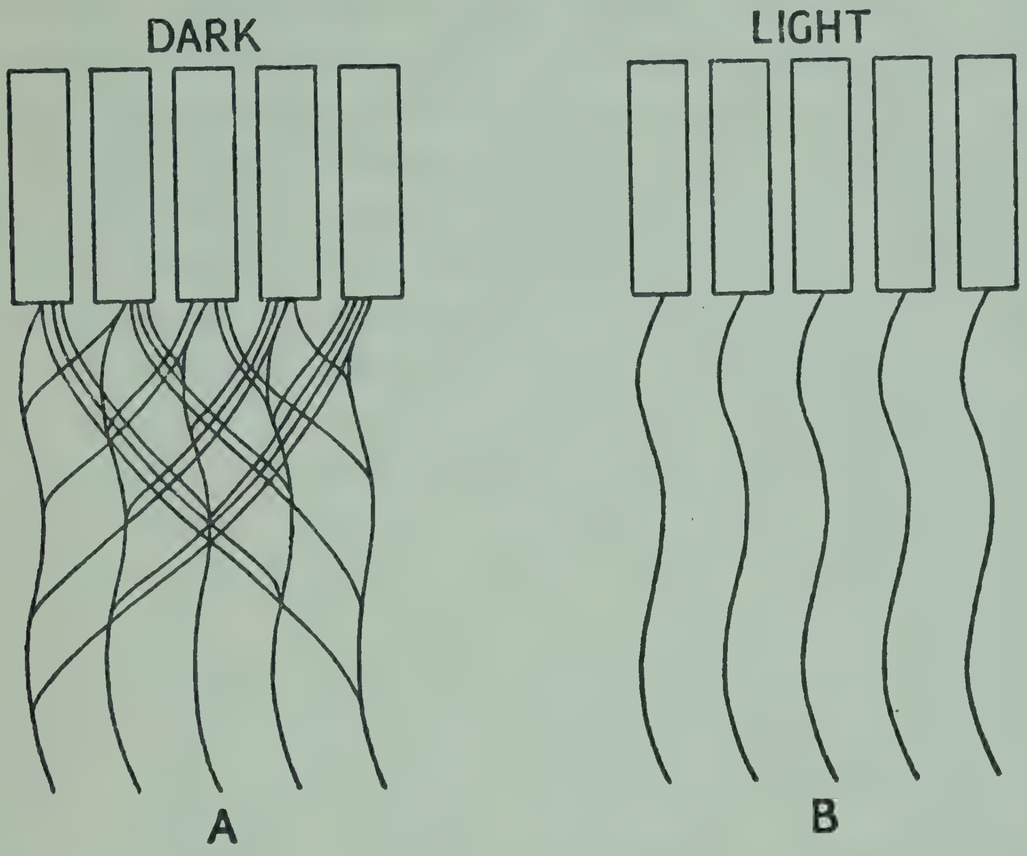


FIG. 39, 24

Illustrates the suggestion discussed in the text that there is more overlapping of the nerve supply to the rods and cones during dark adaptation than during light adaptation. The numerous interconnexions in A are not to be thought of as new nerve fibres but as synaptic interconnexions which function only in the dark (R. J. Lythgoe (1940), The mechanism of dark adaptation. *Brit. J. Ophthalmol.* **24,** 40).

We may conclude that the classical theory which accounts for changes in rod threshold during light and dark adaptation in terms of the rhodopsin content of the rods is inadequate. Nervous interaction in the retina is probably also involved in sensitivity changes.

BINOCULAR VISION

Movements of the Eyes

The eyeballs are supported by the fatty tissue of the orbit with the fascial sheath of the eyeball forming a sort of joint capsule or socket. This is not to be regarded as a rigid socket like the acetabulum for the head of the femur since the eyeball can be moved about 4 mm from side to side and about 1 mm up or down. The tendons of the external eye muscles give off a check ligament before piercing the capsule to reach their attachments on the sclera. The lateral and medial recti rotate the eyeballs 'out' or 'in' round a vertical axis, but the other four muscles cause rotation around oblique axes so that they produce movements with vertical and horizontal components. These movements are controlled by the nuclei of the III, IV (superior oblique) and VI (lateral rectus) nerves in the midbrain. The groups of nerve cells responsible are closely linked so that both eyes act together and conjugate movement, convergence and accommodation all occur together in a harmonious fashion. For example, on glancing from a distant object to a near object the visual axes are altered to pass through the near object, the axes converge and accommodation occurs. All this happens without double vision since the images of objects fall on corresponding points on the two retinæ. Corresponding points are distributed around the two foveæ so that those in the right half of one retina correspond with those in the right half of the other; those in the upper half of one retina with those in the upper half of the other, and so on. If the two eyes do not move concomitantly then the two images do not fall on corresponding points and diplopia, or double vision, occurs, provided that both images reach consciousness. In the squints found in children in the presence of refractive errors the image in one eye is neglected or suppressed so that diplopia does not occur. The vision in the eye in which the image is suppressed gradually deteriorates unless its image is brought to consciousness by frequent occlusion of the other eye. When squint occurs in later life diplopia is likely to be very obvious to the patient. The false image is projected as indicated in Fig. 39, 25. The clue to the understanding of the peculiar displacement is that if an image falls on one side (say the left) of the fovea it is projected on to the other side (in this case the right) of the external world.

In reading, the words being read are fixated, then the eyes move quickly over a certain distance to the next fixation point so that four or five fixations occur per line of print according to the individual. This quick movement of about 40° in 0·1 second, often called a saccadic movement, cannot be hastened or slowed. In looking at scenery from a moving vehicle fixation of a point occurs followed by a slow movement of the eyes to keep the point fixated, then a saccadic movement occurs to the next fixation point, and so on. A similar oscillatory movement with the rapid component outward and the slower inward towards the centre of fixation may sometimes be observed in disease of the nervous sytem when the eyes are conjugately deviated to either lateral field, upwards or downwards. This phenomenon is called nystagmus.

Fixation is brought about partly voluntarily and partly involuntarily, that is reflexly, and can be supposed to occur somewhat as follows. The image of an object on the right falls on the left of the retina and impulses pass up to the visual cortex in the left occipital lobe, and from there go forward again to the superior quadrigeminal bodies on the right side and thence to the nuclei of the third and fourth and sixth nerves which control the external eye muscles. The nuclei on the right side of the mid-brain control the muscles which make the eyes move to the right—namely, the right lateral rectus and the left medial rectus. In the latter case fibres cross over the mid-line from the group of nerve cells concerned in the right oculomotor nucleus to pass out in the left oculomotor nerve. The nuclei for the lateral rectus (VI) and the medial rectus (III) are connected by the medial longitudinal bundle so that conjugate deviations of the eyes occur. When the eyes move to the right the image (previously on the left of the fovea) is brought up to the fovea where vision is most distinct, and in this position stimulation of the right and left halves of the retina is equal and so fixation is maintained. The eyes do not stay perfectly still in fixation but show fine jerky movements of a few minutes of arc every second or so. This view of fixation explains the nystagmus which occurs where foveal acuity is poor either because of disease or because of insufficient illumination, as in miner's nystagmus. The eyes may be thought to be groping for the best position for the retinal image. Kittens kept in the dark for

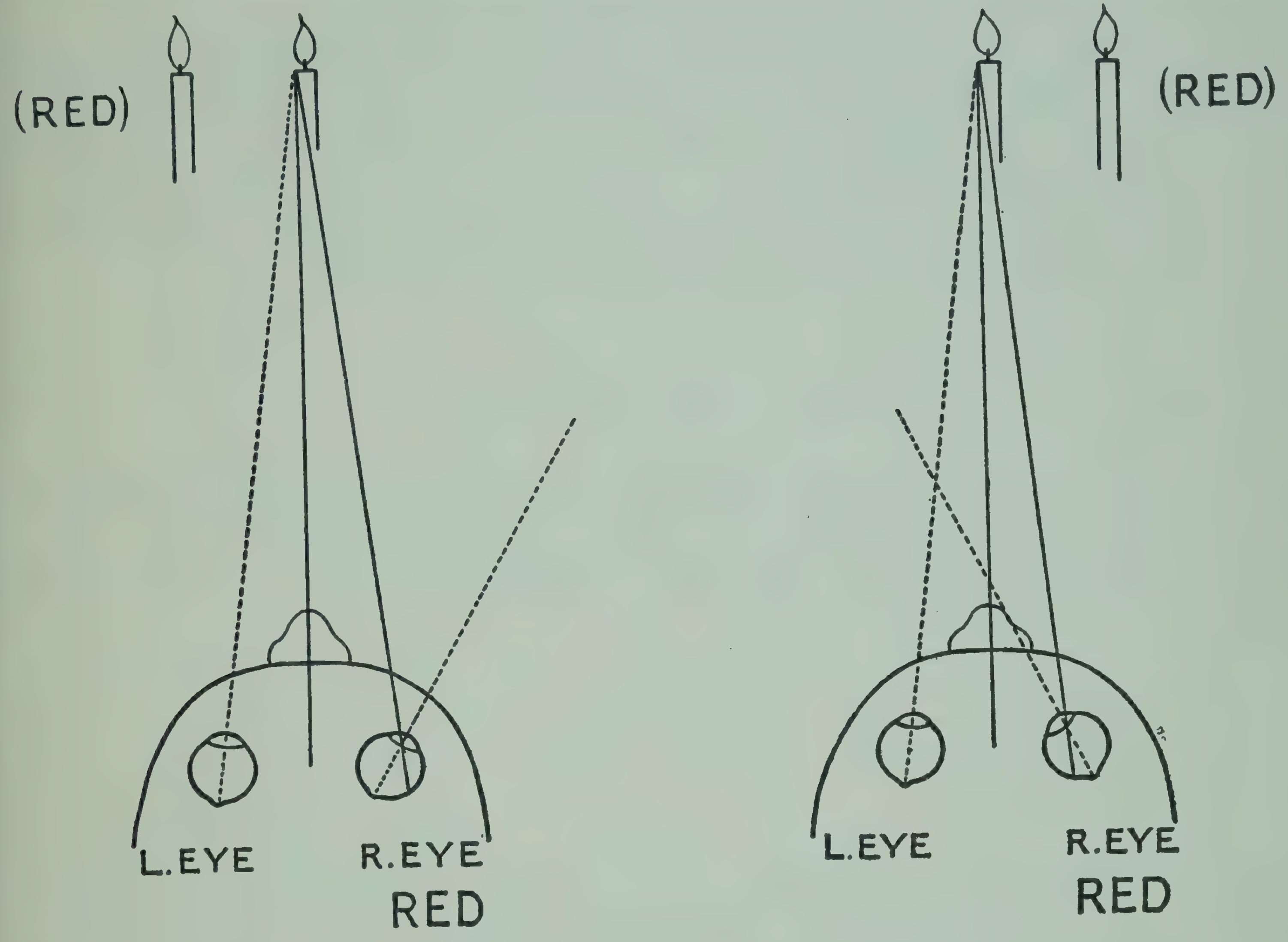


Fig. 39, 25

Diagram to show the projection of the false image from the squinting eye. With a red glass placed in front of the right eye, the image has a red colour. The red image is projected in the direction in which the paralysed muscle would normally turn the eyeball. In the diagram on the left the right medial rectus is paralysed and an external squint is produced. In the other diagram the right lateral rectus is paralysed giving an internal squint. (*After* Maitland Ramsay (1920), *Clinical Ophthalmology for the General Practitioner*, Plate XVIII. London: Henry Frowde, Hodder & Stoughton.)

four to seven weeks develop nystagmus which diminishes when they are brought into the daylight.

Voluntary eye movements are under control of nerve cells lying in area 8$\alpha\beta\delta$ which is placed rostrally to the precentral motor cortex (Fig. 49, 5). Their axons pass down with the rest of the cortico-spinal fibres to the superior quadrigeminal bodies where they cross to the oculomotor nuclei on the opposite side. In this way impulses from the motor cortex of one side are able to turn both eyes to the other side. Removal of area 8$\alpha\beta\delta$ causes loss of voluntary eye movements.

If a certain part of the cerebral cortex at the middle of the inferior parietal lobule (angular gyrus) is stimulated electrically the eyes move to the other side. The fibres from this gyrus must, therefore, cross over to reach the nuclei in the mid-brain. Since the angular gyrus is connected to the motor cortex and to the occipital cortex this gyrus may be an area for integration of reflex and voluntary eye movements.

Perception of Depth

With one eye the distances between two objects and between an object and the observer can be estimated with a fair degree of accuracy. Even a still photograph gives such clues. For example, the distance to an object of known size, a man or a motor-car, is indicated by the *size* of the retinal image. *Perspective* with convergence of parallel lines gives a clue to distance in street scenes. The shape of the corner of a room is recognized by the *light and shade*. A near object may obscure one further away (perspectival occlusion); a man's face cannot be seen behind his newspaper. If a small amount of movement of the head is allowed, or if we look at moving pictures, an important additional clue, *parallax*, becomes available, due to the relative movement on one another of objects in the field of view. If the observer goes through a wood looking at the trees in the middle

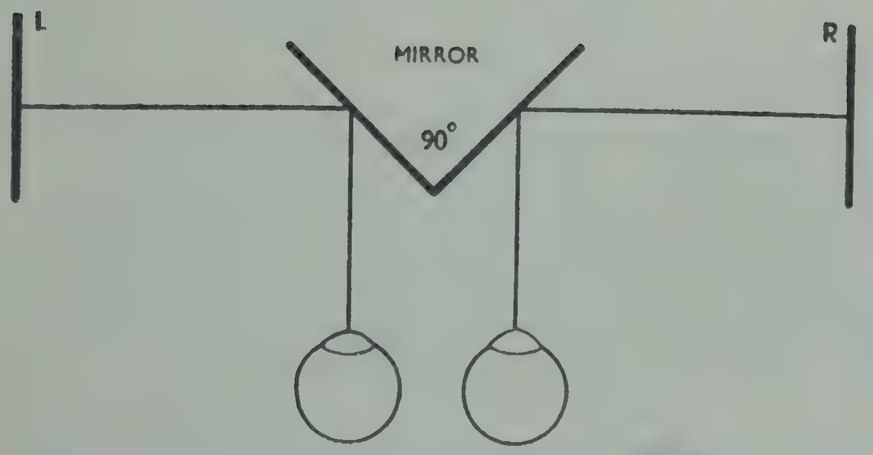


Fig. 39, 26, A

Wheatstone's stereoscope.

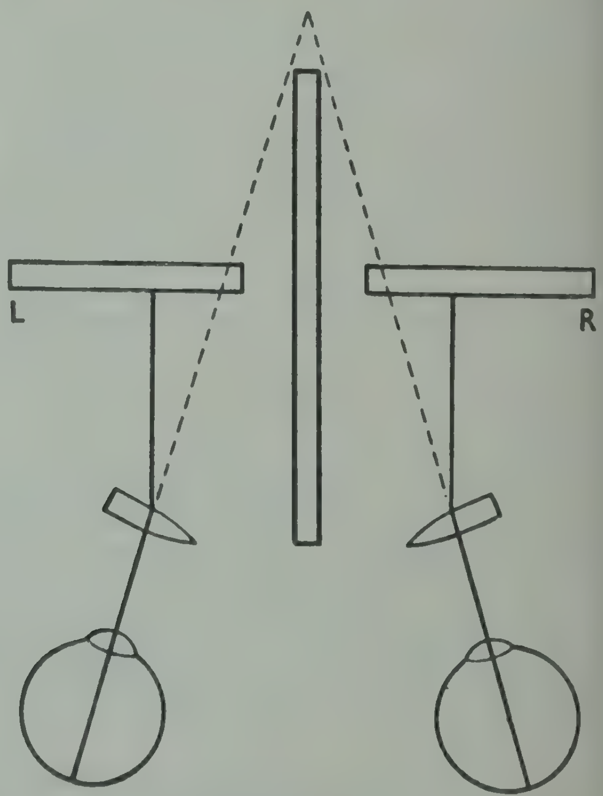


Fig. 39, 26, B

Brewster's stereoscope.

distance they appear stationary, but the trees near at hand pass back, while the trees in the distance move forward. In other words, the external world appears to revolve around the point of fixation. All these clues as to the distance of objects are monocular and do not require two eyes.

If near objects are examined with two eyes *stereoscopic vision* is possible. In binocular vision all the clues of depth and distance already listed are present and in addition a three-dimensional picture of the external world is built up on the basis of the slight disparity between the two retinal images which are fused by some mental process. In addition it can be shown experimentally that the convergence of the eyes necessary to view a near object is a factor in judging its distance. By closing each eye in turn while looking at an object, such as a cube held in the hand, the difference between the two images can be easily appreciated. If two photographs taken from slightly different viewpoints, usually 2½ or 3 inches apart, are presented to the eyes so that the left-hand view reaches the left eye and the right-hand view the right eye only, then the picture appears in relief and the relative positions of objects can be clearly appreciated. The two main forms of the stereoscope or apparatus for presenting two such photographs to an observer are shown in Fig. 39, 26. The Wheatstone apparatus is used to examine X-ray photographs taken from slightly different viewpoints for location of foreign bodies in the tissues. The examination of pairs of aerial photographs by this method shows the rise and fall of the ground clearly and makes interpretation of the photographs much easier. Stereoscopic (3 D) moving pictures are now commonplace but the practical difficulties are considerable because the condition to be satisfied to obtain true stereoscopy, namely that two slightly dissimilar images shall fall simultaneously on the two eyes, cannot be comfortably attained at present.

Under ordinary conditions three dimensional vision arises from binocular perception aided by the monocular clues already mentioned such as parallax, perspective and size. Stereopsis is, therefore, as Wright says "the result of a mental analysis and synthesis of quite a complicated array of visual data."

REFERENCES

Adler, F. H. (1953). *Physiology of the Eye: Clinical Applications.* 2nd ed. London: Henry Kimpton.

Bell, G. H. (1948). Visual physiology and the ciné-film, pp. 20–32 of *Science in Films,* I. Ed. B. Lloyd. London: Sampson Low, Marston & Co.

Carmichael, L. & Dearborn, W. F. (1948). *Reading and Visual Fatigue.* London: Harrap.

Davson, H. (1949). *The Physiology of the Eye.* London: Churchill.

Duke-Elder, W. S. (1932). *Textbook of Ophthalmology.* London: Kimpton.

Fincham, E. F. (1951). The accommodation reflex and its stimulus. *Brit. J. Ophthal.* **35,** 381–393.

Granit, R. (1947). *Sensory Mechanisms of the Retina.* Oxford University Press.

Hartline, H. K. (1941–1942). The neural mechanisms of vision. *Harvey Lect.* Ser. **37,** 39–68.

Helmholtz, H. (1896). *Handbuch der physiologischen Optik.* Hamburg: Voss. Translation from the third (posthumous) edition, Optical Society of America (1924).

Mutch, R. (1944). The lacrimation reflex. *Brit. J. Ophthal.* **28,** 318–336.

Physiology of vision (1953). *Brit. med. Bull.* **9,** 1–80.

Pirenne, M. H. (1948). *Vision and the Eye.* London: Pilot Press.

Pirie, A. & Van Heyningen, R. (1956). *Biochemistry of the Eye.* Oxford: Blackwell.

THOMSON, L. C. (1952). A review of ideas on colour perception. *Ophthal. Lit.* **6,** 3–40.

WALD, G. (1953). Vision. *Fed. Proc.* **12,** 606–611.

WESTON, H. C. (1949). *Sight, Light and Efficiency.* London: Lewis.

WRIGHT, W. D. (1938). *Perception of Light.* London: Blackie & Son.

WRIGHT, W. D. (1946). *Researches on Normal and Defective Colour Vision.* London: Henry Kimpton.

STEVENS, S. S. (Ed.) (1950). *Handbook of Experimental Psychology*, Chaps. 22–24. London: Chapman & Hall.

CHAPTER 40

SPEECH AND HEARING

HEARING is one of the perceptual processes by which animals are continually being informed about their environment. In man the general information from the ears, such as the proximity of a source of sound, is augmented by the more specific information contained in speech. Speech and hearing must be regarded as two complementary activities which subserve the function of communication. In radio-communication the transmitter and the receiver may also be regarded as a complementary pair—if one is out of action the other is useless. If severe deafness occurs in infancy before vocalization has reached the stage of speech the child fails to develop the ability to speak, although he may be taught to speak intelligibly by special methods. In fact, however, unless his hearing is normal, or nearly so, it is difficult for anyone to acquire natural or normal speech.

While there is great variation among children in the rate and manner of acquiring speech it is important to have some general idea of the way in which it is developed. The baby's first cries usually express hunger or discomfort, but if they always have the effect of attracting attention he may learn to cry for this reason alone. In three months he acquires several vowels and a few consonants and learns to smile. In the next six months he imitates spoken sounds but without any realization of their significance and at this stage his smiles may develop into laughter. The first meaningful use of words, perhaps connecting a sound with a toy, or with himself, occurs when he is about one year old. The time at which this is accomplished depends on many factors, such as the state of health and level of intelligence, and it may be accelerated or delayed by learning to walk. In the second year the child often uses one-word sentences, such as "Up," meaning "Lift me up." At the end of the second year two-word sentences appear and by the age of five years sentences of five words are being used. Thereafter the vocabulary increases rapidly.

THE ORGANS OF SPEECH

So far as speaking is concerned, the lungs and chest wall act as bellows which drive air through the larynx between the two sharp folds, the *vocal folds* or vocal cords, and set them into vibration, much in the way that air blown through the lips makes them vibrate. Normally speech is produced on expiration although an abnormal type of speech can be produced on inspiration. The vocal folds stretch from the thyroid cartilage in front to the mobile arytenoid cartilages at the back of the larynx; the triangular space between them is called the *rima glottidis* or glottis. The muscles controlling the arytenoids (Fig. 40, 1) open or close the rima glottidis, while the tension of the folds is regulated by the vocalis muscle lying in each fold and by the cricothyroid muscle which tilts the thyroid cartilage and so elongates the vocal folds (Fig. 40, 2). When the interior of the larynx is inspected through a mirror placed obliquely against the soft palate, the posterior ends of the vocal folds are found to be about 8 mm apart in quiet respiration; they open a little further on inspiration and come together a little on expiration. When the subject is asked to say "ah" or "ee" (that is, to phonate) the folds are seen to approximate. The frequency of a note can be regulated by the tension in the vocal folds.

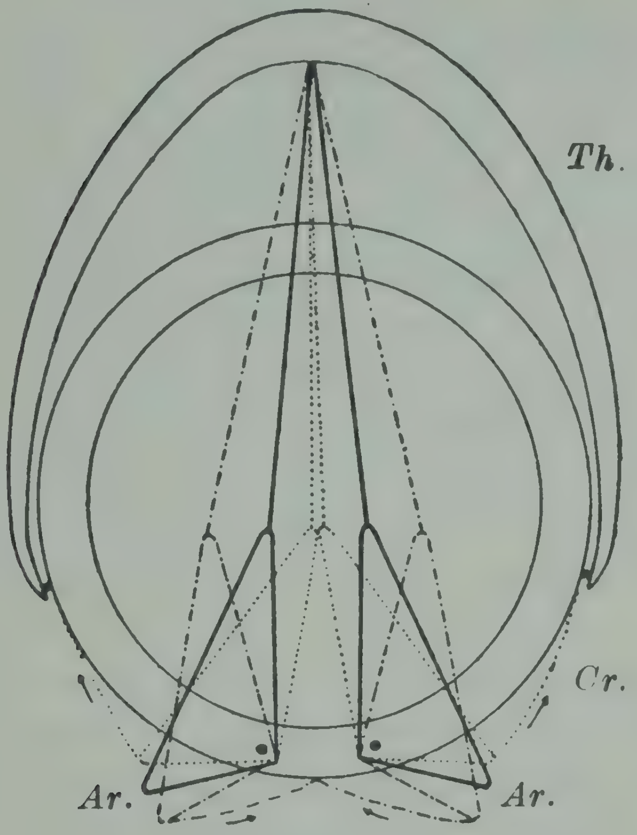


Fig. 40, 1

Cross-section of the larynx, to show the cricoid, *Cr.*; thyroid, *Th.*; arytenoid cartilages, *Ar.* The continuous line shows the parts at rest, the dotted line under the action of the lateral crico-arytenoid muscle, and the dot-dash line under the action of the posterior crico-arytenoid. The arrows show the direction of the pull of these muscles. The transverse and oblique arytenoid muscles (not shown) bridge the gap between the arytenoids; when they contract the arytenoid cartilages come together and the glottis is narrowed (D. Noel Paton (1920), *Essentials of Human Physiology*, 5th ed., p. 564 Edinburgh: Green & Son).

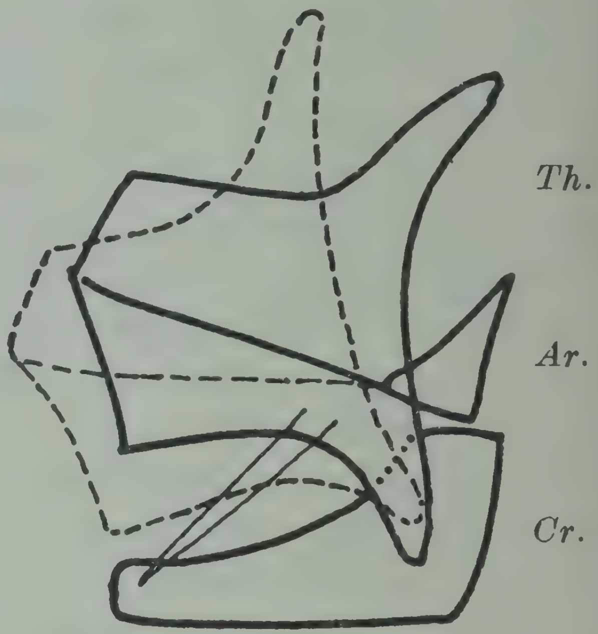


Fig. 40, 2

Side view of the cartilages of the larynx. *Cr.*, cricoid cartilage; *Ar.*, left arytenoid cartilage; *Th.*, thyroid cartilage. The dotted line shows the change in the position of the thyroid by the action of the crico-thyroid muscle, and the stretching of the vocal folds which results (D. Noel Paton (1920), *Essentials of Human Physiology*, 5th ed., p. 564. Edinburgh: Green & Son).

If the tension increases, the frequency rises and there may be considerable increase, up to 50 per cent., in the length of the folds. The vocal folds are lubricated by thin mucus secreted mainly in the ventricles, which are spaces formed by folds of the laryngeal mucosa immediately above the vocal folds. In whispering the anterior two-thirds of the vocal folds are approximated, free escape of air occurring in the space between the two arytenoid cartilages posteriorly.

The sound produced in the larynx, the laryngeal tone, is greatly modified by the acoustic properties of the mouth, throat and nasal cavities. The tongue, by alterations in its shape and position, has the main control over the resonant characteristics of the oral cavity (Fig. 40, 3) but the positions of the lips and jaws are also of importance in the production of speech sounds. The vowel sounds are produced by vibrations of the vocal folds (that is, are voiced), the air stream passing freely through the mouth. In the production of a consonant the air stream is either partially or completely obstructed so that it cannot issue freely from the mouth. Some of the consonants are very short, for example *t* and *p* and *k*, and may be regarded as particular ways of beginning or ending vowel sounds, but some consonants, such as *m*, can be spoken without a vowel. The larynx does not appear to be involved in the production of many consonants, that is the majority are unvoiced. A laryngeal component can, however, be easily added to some consonants to make them voiced; in this way *f* becomes *v*, and *s* becomes *z*.

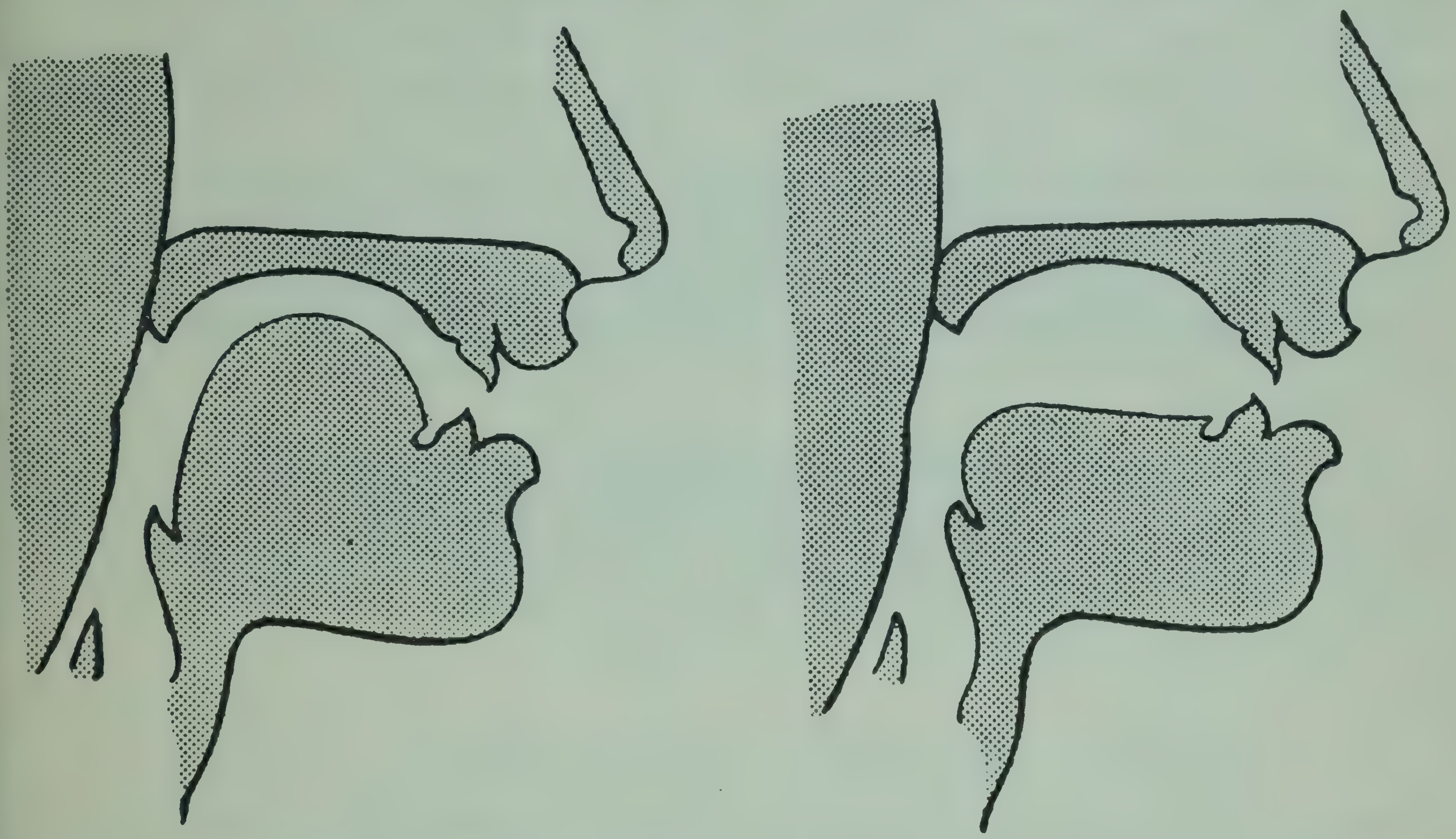


FIG. 40, 3

Tongue position of vowel i.
i (ee *as in* see)

Tongue position of vowel a.
a (*as in* calm)

The positions of the tongue used in the formation of two vowels to show how the shape of the cavity of the mouth is altered. In this way the resonant properties of the cavity can be made to vary over a wide range. (*After* I. C. Ward (1931), *The Phonetics of English.* Cambridge: Heffer.)

The vagus nerve carries motor and sensory fibres to and from the larynx. The superior laryngeal branch after receiving a twig from the sympathetic divides into the internal laryngeal nerve which is sensory to the upper larynx, and the external laryngeal nerve which supplies the cricothyroid muscle. All the other intrinsic muscles are supplied by the recurrent laryngeal nerve which also gives sensory fibres to the larynx below the vocal folds. Section of one recurrent nerve causes hoarseness, since the tension of the vocal folds is reduced and apposition in phonation is difficult; in bilateral recurrent paralysis the glottis remains open, the vocal folds are motionless and the voice is reduced to a whisper. If the superior laryngeal nerve is cut, there is anæsthesia of the upper part of the larynx and the voice is hoarse as the vocal folds cannot be made tense.

There are many good reasons for regarding the upper resonators (mouth, throat and nose) and not the larynx as the main organs of speech, but perhaps the most dramatic evidence is furnished by the phenomenon of œsophageal speech. After complete removal of the larynx the trachea is made to open on the surface of the neck while the œsophagus retains its relation with the mouth. If the patient acquires the trick of swallowing air he can bring it up in a belch which sets the lower edge of the inferior constrictor of the pharynx (cricopharyngeus muscle) into irregular vibrations which are conveyed to the mouth cavity. The patient uses the same mouth positions as he did before his operation to modify this œsophageal sound and although the vibrations of this pseudolarynx are extremely irregular compared with those of the larynx, yet he can with practice produce œsophageal speech which, although rather hoarse, is readily intelligible. For natural and normal speech the larynx is of course essential, but, since it is the modulation of the laryngeal (or pseudolaryngeal) tone which produces the various speech sounds and conveys the information from speaker to listener, it can with reason be asserted that upper resonators of the mouth and throat are in fact the essential organs of speech. They are put into the proper position for the formation of speech by the help of the kinæsthetic sense which gives information to the brain of the position of the tongue, lips and jaws. In addition the ear, in both normal and œsophageal speech, acts as a proprioceptor and muscular adjustments are made to produce the desired speech sound. If this proprioceptor does not function well, that is if there is deafness, these adjustments cannot be made accurately and the speech is defective.

Analysis of Speech Sounds

It is well known now that the sounds of speech and of musical instruments are very complex but are capable of analysis by physical and mathematical methods into a fundamental tone and a number of harmonics or overtones of higher frequency. In the middle of last century Helmholtz was able to make an analysis of such sounds using his own ear aided by resonators to obtain greater sensitivity. He could, however, pick out only the loudest harmonics and was unable to detect the weak components which are often important characteristics of speech sounds. The energy involved in speech is so small (10 to 25 microwatts in ordinary conversation) that without amplification no device is sufficiently sensitive to make an accurate record. The full complexities of speech and musical sounds were first revealed by using a microphone and valve amplifier connected to a galvanometer which records by a beam of light on a moving photographic film. Typical records of speech wave forms are given in Fig. 40, 4, in which the abscissa is time and the ordinate can be regarded as air pressure at the speaker's mouth. Each sound, consonant or vowel, has its own waveform. The vowels have a waveform which recurs regularly throughout their duration at the frequency of vibration of the vocal folds. The pattern is usually very complicated, showing that there is, in addition to the fundamental laryngeal tone, a considerable number of harmonics of much higher frequency, some of which may even be of greater amplitude than the fundamental. Consonants like *p*, *b*, *t*, *d* are so short—electrical engineers refer to them as 'transients'—that no sensation of pitch is given by them; sibilants, *s* and *z*, may last as long as vowels and are characterized by small waves of very high frequency. Transitional consonants like *m*, *n*, *ng* also last for some time. Voiced consonants such as *z* and *v* have a waveform which recurs at the fundamental laryngeal frequency. A study of oscillograms of vowels suggests that each puff of air that comes up from the larynx sets the cavities of the mouth and throat into vibration. This vibration gradually diminishes until the next puff of air comes up from the larynx. This is well illustrated by the oscillogram of the vowel "ah" as in father, when spoken by a man (Fig. 40, 4, A). When the same vowel is spoken by a woman (Fig. 40, 4, B) the time between the laryngeal puffs is reduced to about half with the result that the vibrations of the air in the mouth diminish very little before the next puff comes along. Although the two waveforms of "ah" in Fig. 40, 4, are quite different, the listener has no difficulty in recognizing that both speakers are pronouncing the same vowel. It is thus unlikely that the ear recognizes vowels by their waveforms and this is confirmed by the following considerations. When one listens to speech in a room, part of the sound arrives directly from the speaker to the ear and part indirectly after reflexion from the walls; the reflected waves are out of phase with the direct wave. In many gramophone or microphone amplifiers the higher frequencies are not emitted from the loudspeaker in their original phase relationship to the low frequencies. In both of these examples the original waveform is considerably changed, but the sound does not appear to be altered.

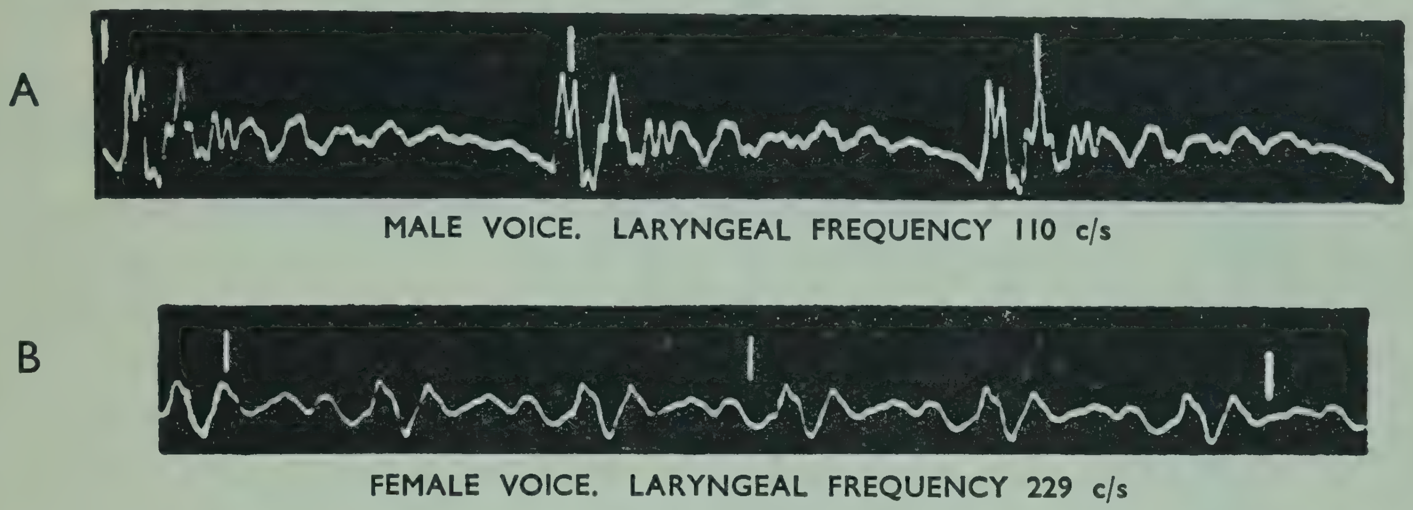


Fig. 40, 4

Oscillograms of the vowel sound "ah" as in "father." (A) Spoken by a man and (B) spoken by a woman. The vertical white lines are 0·01 sec apart. (Harvey Fletcher (1929), *Speech and Hearing*, Fig. 37, p. 49. London: Macmillan). (Bell Telephone Laboratories.)

An acoustic spectrum is obtained by analysing a speech waveform to obtain the amplitude and frequency of the various components and some examples are given in Fig. 40, 5. Examination of such spectra gives an indication of the clues which the ear uses to distinguish speech sounds. If spectra of the same vowel spoken at different pitches (or, to express it more correctly, at different laryngeal frequencies) are compared it will be found that, in spite of the alteration in the frequency of the fundamental as the pitch of the voice is raised, the overtones are relatively unchanged and a characteristic grouping of the higher frequency components can be recognized in all the spectra. Presumably it is this grouping or 'formant' which is recognized by the ear. Once again it is clear that it is the higher frequencies produced in the mouth and throat which are important for differentiating and recognizing the vowels. Spectra of whole words have also been found to have recognizable similarities in spite of individual differences in pitch and quality. By the use of somewhat elaborate apparatus such spectra can be displayed on a cathode-ray oscillograph as the sounds are being spoken. Whether this so-called 'visible speech' would be useful to deaf persons remains to be seen.

The rate of vibration of the vocal folds, which can be easily found from oscillograms, gives a clue to the sex of the speaker. A deep-voiced man may have a laryngeal tone of 90 c/s (cycles per second), but in the average man it is usually between 125 and 145 c/s. The average woman's laryngeal tone is from 230 to 256 c/s, but in a shrill-voiced woman it may be as high as 300 c/s. The deepest bass note that can be sung is about 66 c/s, while the highest soprano note is about 1056 c/s. This is a range of four octaves.

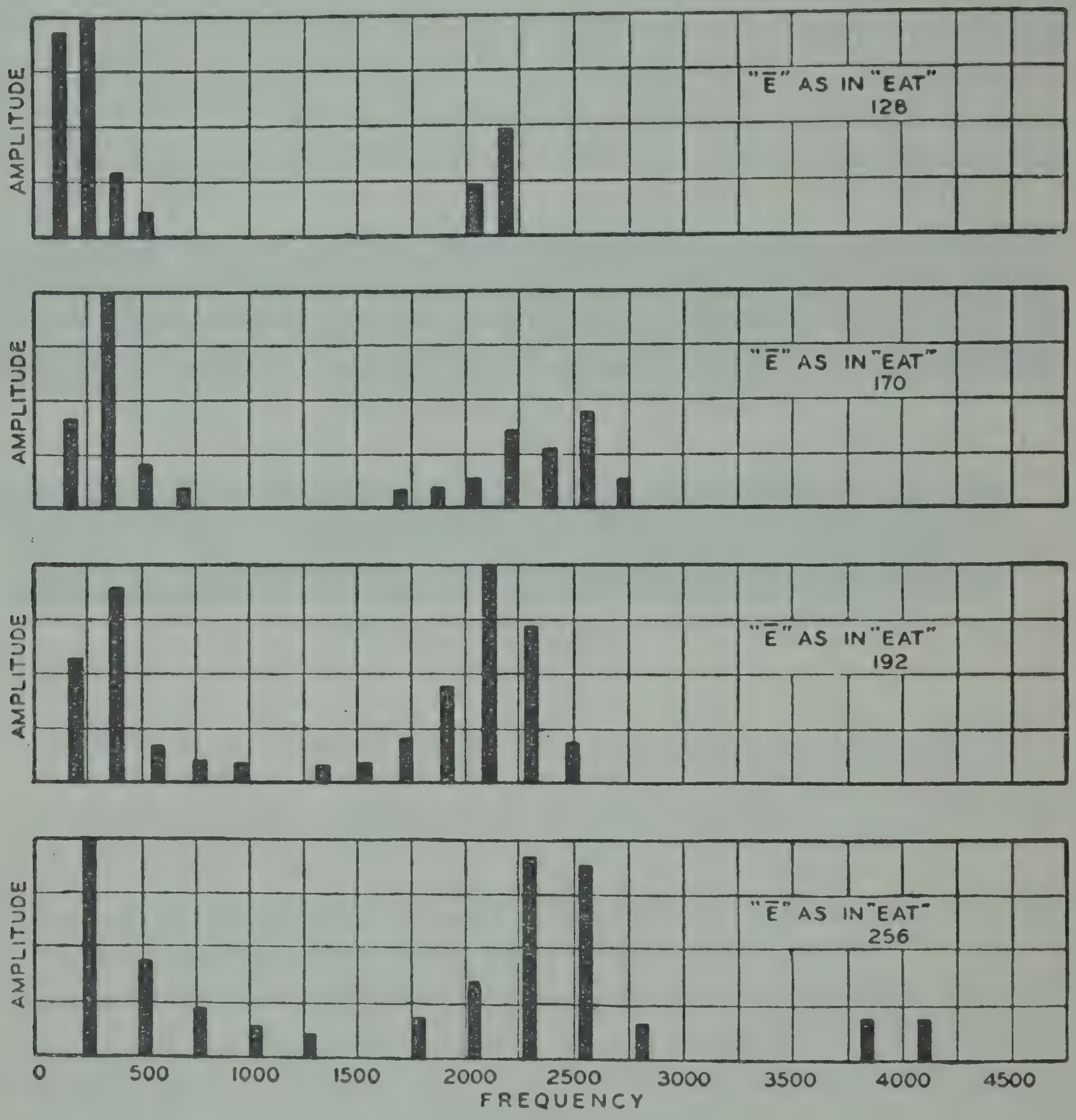


Fig. 40, 5

Acoustic spectra of vowel sound "EE" at different frequencies of the fundamental laryngeal tone (128, 170, 192 and 256 c/s) to show the general similarity of the spectra (Harvey Fletcher (1929), *Speech and Hearing*, Fig. 41, p. 54. London: Macmillan). (Bell Telephone Laboratories.)

The Decibel Notation.—The energy involved in speech can be calculated from oscillograms or measured by suitable apparatus. The range is very great—from 0·001 microwatts in a very soft whisper up to 1000 microwatts in very loud talking. Since the ear, like the other sense organs, follows approximately the Weber-Fechner law which implies that equal steps on a logarithmic scale of intensity are equal steps on a loudness scale it is more convenient to describe differences in power-level in the decibel notation:

Difference in power-level in decibels $= 10 \log_{10}$ (ratio of the two powers).

If the figures quoted above for speech are taken the difference in power-level between a soft whisper and loud talking is

$$10 \log_{10}\frac{1000}{0{\cdot}001} = 10 \log_{10}\frac{10^3}{10^{-3}} = 10 \log_{10} 10^6 = 10 \times 6 = 60 \text{ decibels, or } 60 \text{ db.}$$

Calculation (using this formula) shows that 1 db is equivalent to a 26 per cent. increase in power; 10 db to a ten times increase in power; 20 db to one hundred times increase in power.

The External Ear and Middle Ear

The mobile trumpet-like external ear (auricle) possessed by some of the lower animals may be helpful in localizing the source of a sound and in some animals the auricle twitches when loud sounds are heard. This can be used as a test of hearing. The auricle is not usually mobile in man. From the auricle, the **S**-shaped external auditory meatus, about 2·5 cm long in the adult, passes down to the tympanic membrane. The skin of the outer half of the meatus possesses numerous ceruminous or wax-secreting glands which are modified sweat glands. Accumulation of wax in the external auditory meatus is a common cause of deafness.

The tympanic membrane, about 1 cm in diameter, is composed of radial and circular collagenous fibres covered externally by stratified squamous epithelium continuous with the skin and internally by a mucous membrane of ciliated columnar epithelium. The outer surface of the membrane is concave; the inner convex surface, looking into the middle ear or tympanic cavity, has the manubrium of the malleus (handle of the hammer) attached to it. The malleus articulates with the incus (anvil) and the incus with the stapes (stirrup), the base of which is attached by fibrous tissue to the margins of the fenestra vestibuli (oval window). These very light ossicles in the air-filled tympanic cavity link the ear drum to the inner ear (Fig. 40, 9). The tensor tympani muscle, by its attachment to the root of the manubrium of the malleus, draws the tympanic membrane in, and at the same time, through the articulations between the three ossicles, pushes the stapes into the internal ear (Fig. 40, 10). The small stapedius muscle tends to have the opposite effect, namely to pull the stapes out of the internal ear. In response to sound, or to mechanical stimulation of the external auditory meatus, auricle or surrounding skin, these muscles contract reflexly, pulling in the tympanic membrane and tightening it and pushing the footplate of the stapes inwards, presumably because the tensor tympani is more powerful than the stapedius. The result is that sound transmission across the middle ear is impaired for tones of low frequency and the ear is protected from damage by loud sounds.

The ear drum and the ossicles act as a kind of transformer to convey vibrations of the light medium, air, to the dense medium, water, in the internal ear. The area of the ear drum is about 90 sq. mm, whereas the area of the base of the stapes is only 3·2 sq. mm. The pressure on the fluid under the stapes is thus many times greater than the air pressure which makes the ear drum move. In the course of the condition called otosclerosis, a familial and progressive deafness beginning in young people, the base of the stapes becomes joined by bone to the margins of the fenestra vestibuli. This condition causes severe deafness (Fig. 40, 17).

The pharyngotympanic tube leads from the anterior part of the middle ear to the nasal part of the pharynx. The pharyngeal end is normally closed but during swallowing it opens and allows the air pressure in the middle ear to be equalized with the atmospheric. If the pressure is not equal on the two sides the drum is pushed in (as in the case of the descent of a diver) or drawn out (as on ascending in an aeroplane). Such abnormal displacement produces pain and deafness and damage to the ear drum, but this discomfort can be avoided by frequent swallowing while the air pressure is altering. Deafness may also result when the tube is blocked by the swelling of the mucous membrane which accompanies a 'cold' in the head; in this case the oxygen in the middle ear is absorbed and the ear drum is pushed in by the external atmospheric pressure.

The conical tympanic membrane is normally kept tense by the tensor tympani and can only move inwards by stretching of its radial fibres. On the other hand, outward movement, since it reduces the tension on the fibres, is relatively easy. This can be proved experimentally by attaching to the tympanic membrane a tiny mirror which reflects a beam of light. When the deflexion of the beam for a given rise of air pressure in the external auditory meatus is compared with the deflexion for the same fall of air pressure it is seen that the former is smaller. Thus when loud sounds arrive at the ear they are passed on to the inner ear considerably distorted, as shown in Fig. 40, 6. The distortion of sounds of low intensity is not appreciable, as any small portion of the characteristic curve is, for practical purposes, straight. If an analysis of the distorted waveform is made it is found to be composed of the original frequencies together with multiples or harmonics of them. The ear thus introduces *harmonic distortion* which, although not readily recognizable with a single pure tone, becomes obvious when two loud tones are sounded together and subjective *summation* and *difference tones* are heard. A referee's whistle actually produces two high-pitched notes (say, 1000 and 1250 c/s), but if it is blown loudly a subjective difference tone of low pitch (250 c/s in this example) is the most prominent component. Summation tones are not so easy to pick out. If two loud tones of 200 and 300 c/s (doh and soh) are sounded, then harmonics of each are produced in the inner ear—200, 400, 600 and 300, 600, 900, and so on, with the following summation and difference tones:

Subjective Summation Tones	Subjective Difference Tones
$200+300=500$ $200+600=800$ $400+300=700$ $400+600=1000$	$300-200=100$ $600-200=400$ $400-300=100$ $600-400=200$

The most prominent are the difference tone of 100 c/s, which is one octave below the 200 c/s note, and the summation tone of 500 c/s, which is one octave and a major third above 200 c/s note.

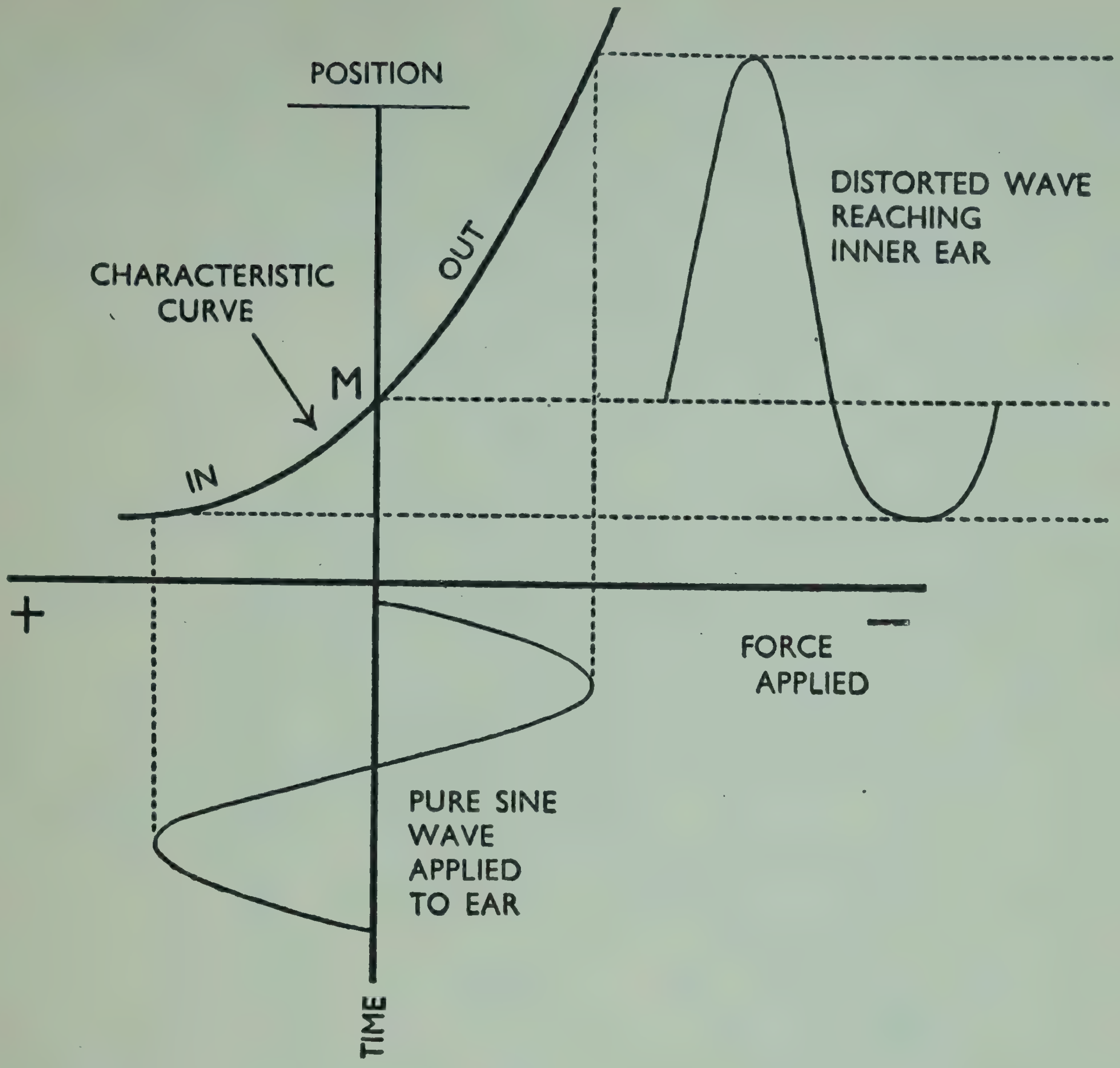

POSITION
CHARACTERISTIC CURVE
OUT
M
IN
DISTORTED WAVE REACHING INNER EAR
+
—
FORCE APPLIED
PURE SINE WAVE APPLIED TO EAR
TIME

Fig. 40, 6

The characteristic curve of the ear. Increase of air pressure, that is the force applied to it, pushes the tympanic membrane in from the mean position M; the reduction of the air pressure by the same amount moves the membrane out by a larger amount. A pure sine wave, if of large amplitude, is thus considerably distorted as shown on the right of the diagram.

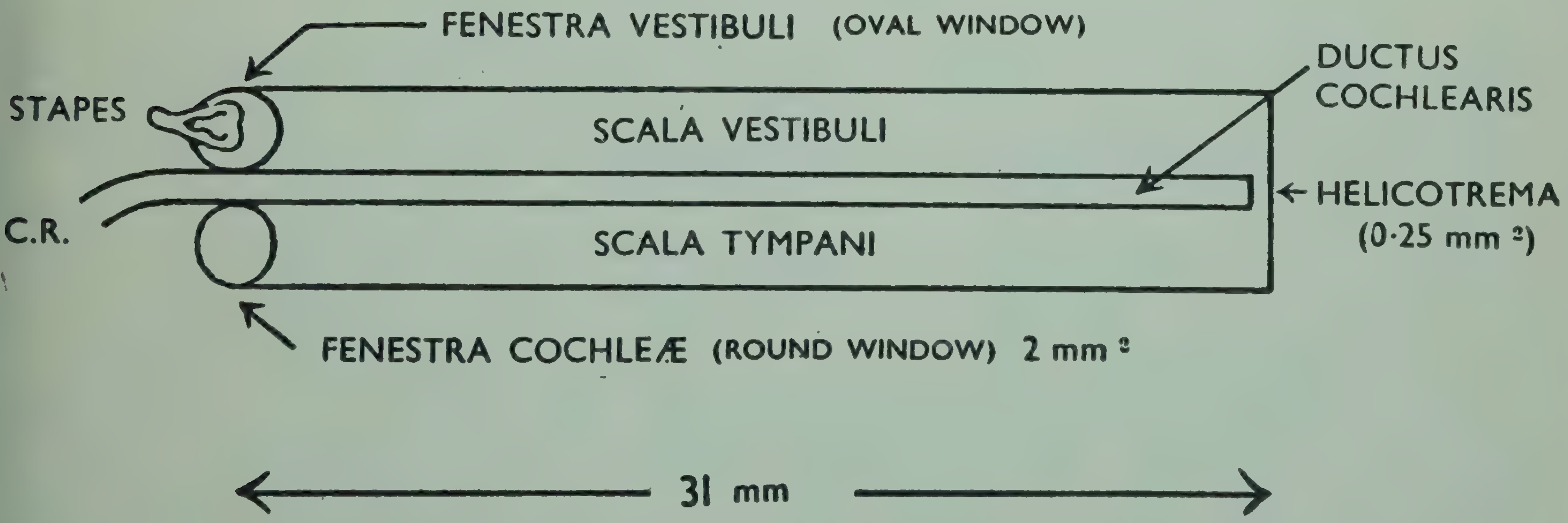

FENESTRA VESTIBULI (OVAL WINDOW)
STAPES
SCALA VESTIBULI
DUCTUS COCHLEARIS
HELICOTREMA (0·25 mm ^{2})
C.R.
SCALA TYMPANI
FENESTRA COCHLEÆ (ROUND WINDOW) 2 mm ^{2}
31 mm

Fig. 40, 7

Basic structure of the inner ear. If this system of three tubes (scala vestibuli, ductus cochlearis and scala tympani) could be wrapped two and three-quarter times round the modiolus (central pillar of the inner ear) the actual anatomical relationships would be reached. C.R. = canalis reuniens which communicates with the saccule (see Fig. 40, 9).

The Internal Ear

In this chapter only cochlear function is described. The semi-circular canals and otolith organs which are concerned with the maintenance of balance and play no part in hearing are discussed on p. 856. A cross-section of the petrous temporal bone is difficult to understand until the basic structure of the cochlea, given in Fig. 40, 7, is appreciated. A diagram of a cross-section through the three canals is given in Fig. 40, 8.

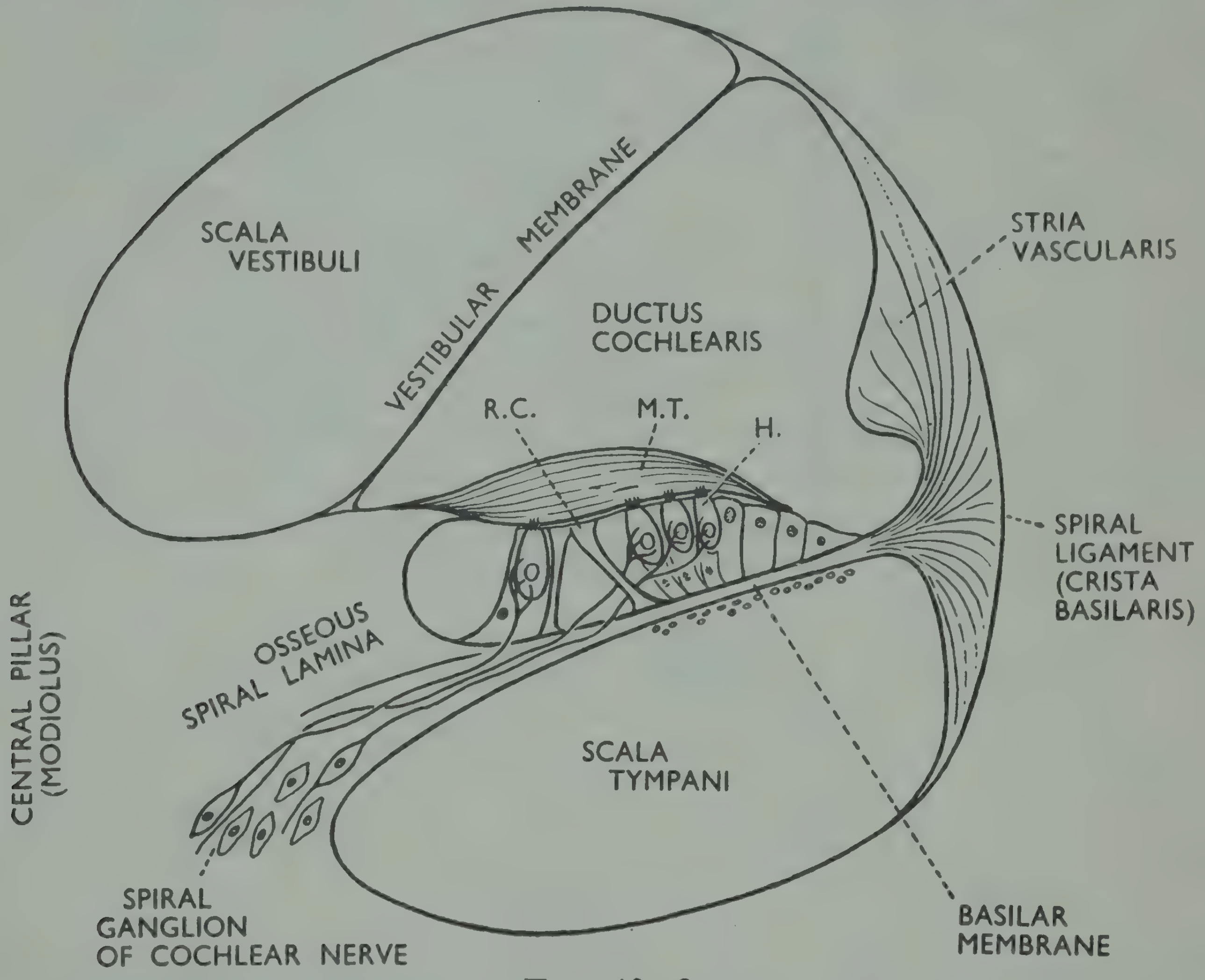

SCALA VESTIBULI
VESTIBULAR MEMBRANE
DUCTUS COCHLEARIS
STRIA VASCULARIS
R.C.
M.T.
H.
SPIRAL LIGAMENT (CRISTA BASILARIS)
CENTRAL PILLAR (MODIOLUS)
OSSEOUS SPIRAL LAMINA
SCALA TYMPANI
SPIRAL GANGLION OF COCHLEAR NERVE
BASILAR MEMBRANE

Fig. 40, 8

A cross-section through the three canals of the cochlea to show the structures on the basilar membrane. H, hair-cells; M.T., membrana tectoria; R.C., rods of Corti. The reticular lamina is a net-like membrane holding in its apertures the free ends of the outer hair-cells. The cochlear partition comprises the basilar membrane, the vestibular membrane and the structures contained in the ductus cochlearis.

If the three canals are wrapped two and three-quarter times round the modiolus (the central pillar of the cochlea) the appearances seen in Fig. 40, 9, can be fully accounted for. An important feature is the *basilar membrane* composed of fibres which run radially from the osseous spiral lamina of the modiolus to the spiral ligament, or crista basilaris, which binds the basilar membrane to the external bony wall of the cochlea. The fibres are short (0·13 mm) near the foramen vestibuli and increase gradually to the apex of the cochlea at the helicotrema, where they are about 0·4 mm long. The spiral organ of Corti is formed of 4000 rods of Corti which make a spiral tunnel on top of the basilar membrane. On either side of this tunnel are hair cells from which filaments of the cochlear nerve leave for the spiral ganglion lying in the modiolus. The fluid in the scala vestibuli and scala tympani (the perilymph) is very like cerebrospinal fluid in chemical composition but may not be continuous with it (Fig. 46, 13). The chemical composition of the endolymph in the ductus cochlearis is somewhat different; it has more potassium and less sodium than the perilymph.

MECHANICAL ASPECTS OF COCHLEAR FUNCTION

If the air pressure in the external ear is increased, the tympanic membrane is pushed inwards, resulting in a movement of the chain of ossicles which tilts the stapes so that it enters more deeply into the vestibular fenestra, so displacing perilymph inwards into the scala vestibuli. If the movement is slow, the fluid is displaced up to the helicotrema, then down the scala tympani, finally bulging the membrane of the fenestra cochleae outwards into the middle ear. The vestibular membrane, being flaccid, can be ignored in these movements. If the movement of the stapes is not slow, but rapid, the inertia of the fluid and resistance to its movement result in pressures being developed which displace the basilar membrane, as indicated diagrammatically in Fig. 40, 10.

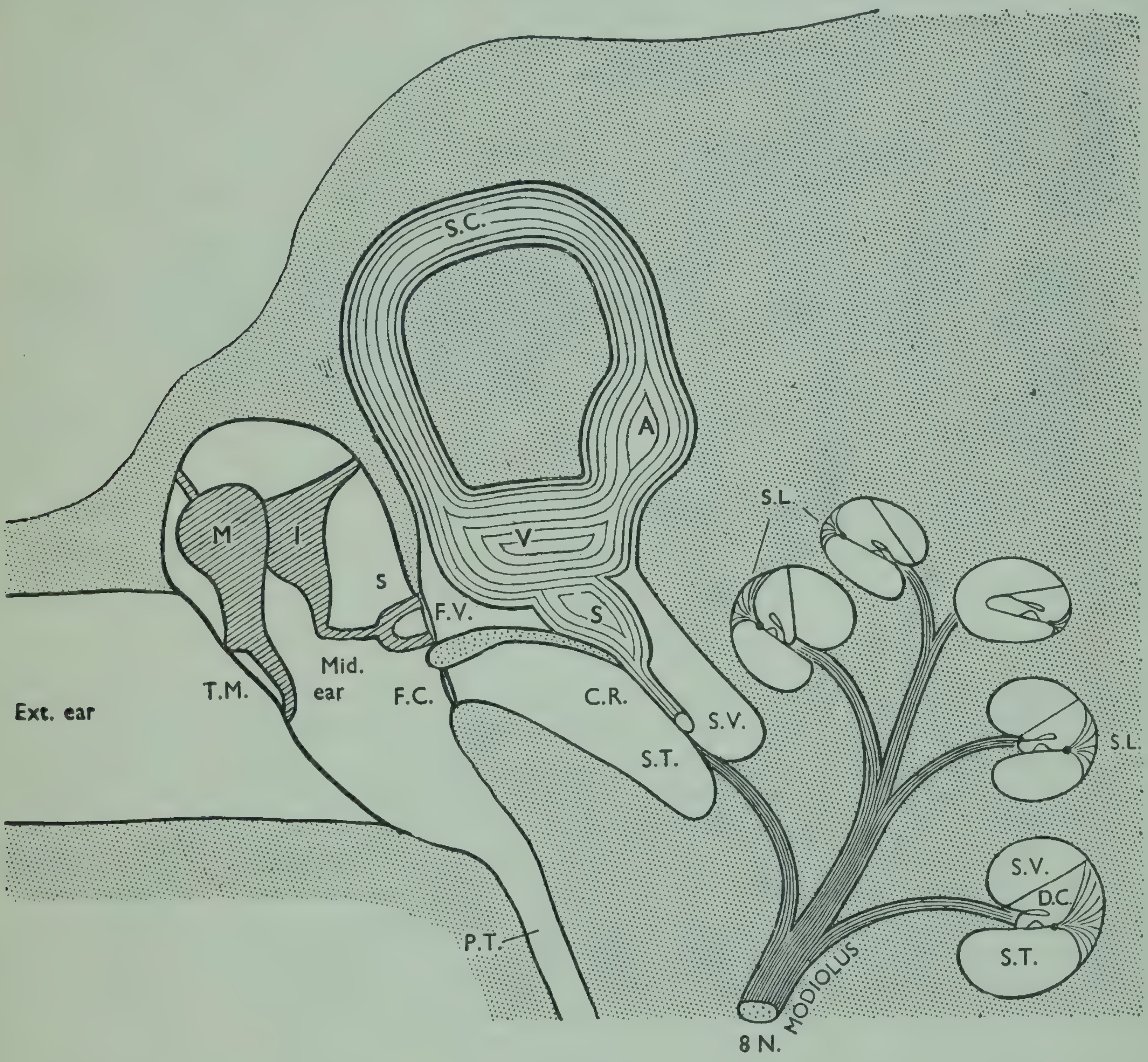


FIG. 40, 9

Diagram of ear. T.M., tympanic membrane; M, malleus; I, incus; S, stapes; SC, semicircular duct; A, ampulla; V, vestibule; S, saccule from which CR, the canal reuniens (cut across), connects with DC, the duct of the cochlea; SV, scala vestibuli; ST, scala tympani; 8N, eighth nerve; SL, spiral ligament or crista basilaris; P.T., pharyngotympanic tube. In constructing the diagram the various structures have been shown in one plane but actually the cochlea lies anterior to the semicircular canals. (*After* McDowall and Keen.)

Various theories of cochlear function have postulated that different portions of the basilar membrane are thrown into vibration by different frequencies of sound in the audible range. The resonance theory of Helmholtz regarded the basilar membrane as a series of tuned resonators, like the strings of the piano, in which the short transverse fibres at the base of the cochlea reacted to high frequency tones, the long fibres at the apex to low frequency tones. Békésy has shown in several ways that these fibres are not in tension. For example, if the basilar membrane is cut the edges of the wound do not gape apart. In consequence the principle of tensioned resonators postulated by Helmholtz has had to be abandoned.

Békésy examined the movement of the human cochlear partition (see legend to Fig. 40, 8) in fresh post-mortem material, in response to vibration applied to the fluid in the scala vestibuli by a small piston which acted as an artificial stapes. The piston occupied the position of the stapes and its movements were electronically controlled and measured. In this way, vibrations of known frequency and amplitude were imparted to the piston, the resulting movements of the cochlear partition being observed microscopically. The amplitude of the vibration showed a maximum which was not very sharp; it was located towards the basal part of the cochlea for high frequencies and towards the apex for low frequencies. The higher the frequency the nearer to the artificial stapes was the maximum amplitude of vibration. Thus some degree of frequency analysis can be performed mechanically by the basilar membrane but not in the sharply selective way needed by Helmholtz's theory. It now appears that the properties of the cochlea in frequency analysis are dependent on the fact that the agitation of the fluid in the cochlea by movements of the stapes causes waves to travel along the cochlear partition. The waves are analogous to those in blood vessels, the necessary elasticity for the occurrence of such waves being supplied by the walls of the cochlear partition. The physical conditions are such that high frequency oscillations die out sooner than low frequency ones, so that the components of a complex travelling wave are eliminated

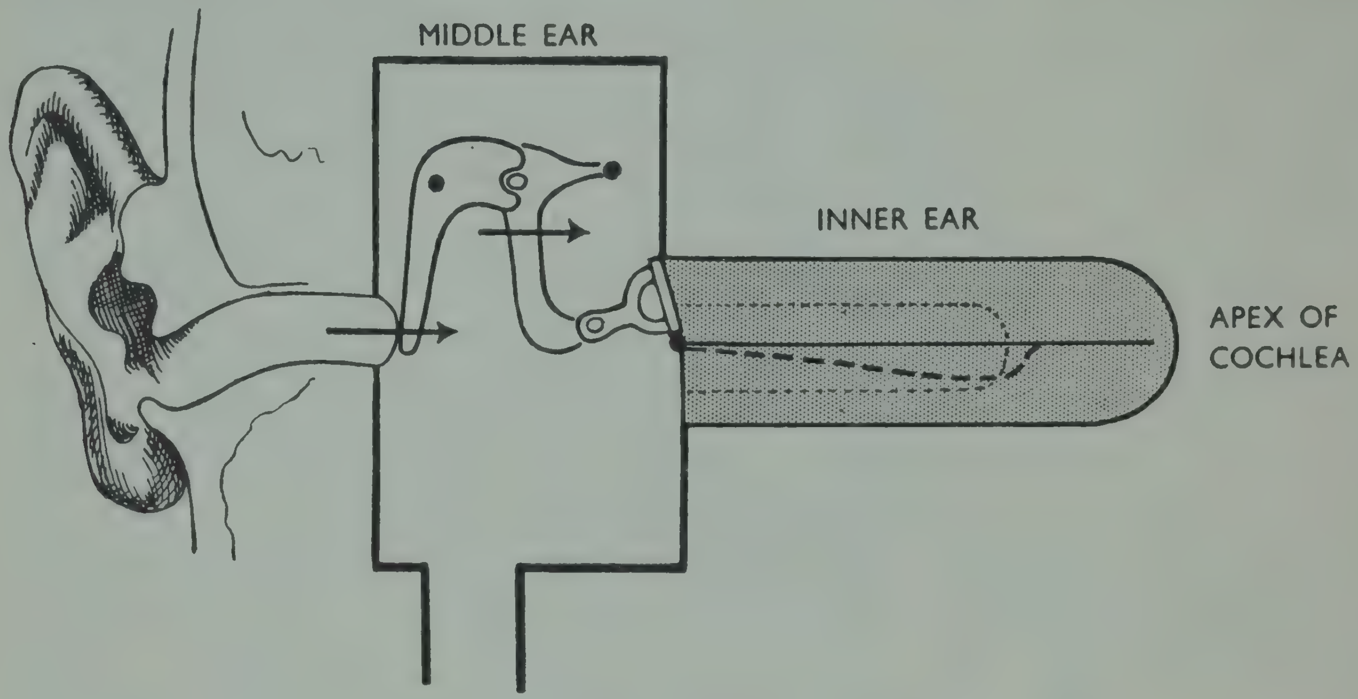


Fig. 40, 10

Displacement of basilar membrane by fluid pressures. The various parts are not drawn to scale. The increase of air pressure in the external ear during the positive phase of a tone causes the basilar membrane to move down and the diminution of air pressure in the negative phase makes the membrane move up. When a pure tone falls on the ear it causes, according to this theory, an oscillatory motion of the fluid in the inner ear along a path such as that indicated by the dotted line. The arrows and dashed line indicate the direction of movement of the ossicles and the position of the basilar membrane during the phase of positive pressure of a sound. The sketch of the chain of ossicles is simplified from the diagram given by R. Guelke & J. A. Keen (1952), *J. Physiol.* **116,** 185. There are three 'fixed pivots' indicated by black circles and two 'free pivots' indicated by open circles.

at different distances from the stapes; in this way high tone localization occurs in the basal region and low tone localization in the apical region. However, the lack of selectivity of this method of frequency analysis makes it clear that mechanical separation of frequencies cannot account for the known powers of pitch discrimination possessed by the ear. For further information on this matter we must look to the electrophysiological evidence of cochlear function.

Electrophysiological Aspects of Cochlear Function

The characteristics of the action of the cochlea have been greatly elucidated by the study of the electrical potentials generated in the inner ear. Three types of electrical effects are found:

1. Steady or direct current (d.c.) potentials (endolymphatic potential);
2. Alternating potentials, known as the cochlear microphonics, and
3. Action potentials derived from activity in the auditory nerve.

Endolymphatic Potential.—In the guineapig cochlea, a potential was found between the endolymph and perilymph by Békésy. By inserting microelectrodes into the perilymphatic and endolymphatic spaces of the cochlear scalæ, he found that the endolymph in the ductus cochlearis, in the absence of any acoustic stimulation of the ear, was electrically positive by some 80 mV to the perilymph in the scala tympani and scala vestibuli. When the microelectrode penetrated into the cellular structures bounding the cochlear duct, negative potentials were usually found. These are derived from the interior of the cells; such internal negativity in cells is to be expected, and is, of course, well known in nerve and muscle cells.

It is also clear from recent work that the boundary of the endolymph space towards the scala tympani is not the basilar membrane, but the reticular lamina and the cellular structures resting on the basilar membrane. The latter, the hair cells and associated fibres are in contact with perilymph, not endolymph, as was previously supposed. The steady positive potential in the ductus cochlearis has been found to be reduced by displacement of the basilar membrane towards the scala vestibuli, and to be increased by displacement of the basilar membrane in the opposite direction, that is towards the scala tympani. These alterations in the potential of the cochlear duct are maintained so long as the deformation of the basilar membrane is maintained. The displacements have been made experimentally by applying pressure to the fluid in the scalæ. It seems probable that the source of the potential is located in the hair cells and in conformity with this is the finding that a radial pull on the tectorial membrane alters the potential. Thus it is likely that a displacement of the basilar membrane alters the potential by changing the position of the hair cells relative to the overlying tectorial membrane so that the forces exerted on the hairs embedded in them are changed.

Cochlear Microphonic.—From the concept of a steady potential capable of modification by the static position of the basilar membrane, it seems logical to presume that, if the basilar membrane were to be vibrated in a direction at right angles to its plane, rapid alternations of a corresponding nature should be noted in the potential in the cochlear duct. Such a supposition provides an explanation for the alternating potentials, known as the cochlear microphonic, produced in the ear when a sound falls on it. The microphonic can be recorded maximally by one electrode in the duct of the cochlea and the other in the scala tympani, that is across the basilar membrane-hair cell complex. It is probable that the hair cells by modifying the steady potential produce the alternating potentials. The microphonic can be recorded from a variety of other situations, such as between scala vestibuli and scala tympani, between the round window and the neck, or from the auditory nerve; in all these situations it is due simply to an electrical spread through the tissues.

The cochlear microphonic corresponds closely to the characteristics of the applied sound stimulus; it is virtually an electrical equivalent of the applied acoustic stimulus, and quite different from the action potentials in the auditory nerve. The cochlea thus acts literally like a microphone; hence the name 'cochlear microphonic.' The potential changes follow the frequency of the applied sound stimulus at all audible frequencies, they resemble quite closely the wave form of the stimulus, they increase in voltage with increase in sound energy; they have an extremely short latent period, and they are not immediately abolished by the death of the animal. The fact that the potentials decline and eventually disappear when oxygen is cut off shows that they depend on the metabolism of the cells in this region. Tasaki, Davis and Legouix inserted small wire electrodes into the scalæ and were able to record microphonics from a length of about 1 mm of the guineapig cochlea. Their electrical recording system was arranged to provide simultaneous cochlear microphonic and auditory nerve traces, each uncontaminated by the other. Confirmation of the existence of travelling waves along the basilar membrane was obtained by the pattern of cochlear microphonics produced at different parts of the cochlea by a sound stimulus; in the passage along the cochlea, tones of low frequency cause waves which reach the apical parts, but the higher the frequency the nearer to the base is the wave extinguished. This work confirms in general that of Békésy in regard to the movement of the cochlear partition. The place of the cochlear microphonics in hearing has been the subject of much discussion since they were first clearly differentiated from the action potentials in the auditory nerve, but it now seems likely that their function is to stimulate electrically the endings of the auditory nerve fibres on the hair cells, so that nerve impulses may be initiated in response to movement of the basilar membrane.

The estimate of the threshold sensitivity of the ear given on p. 725 shows that the amount of mechanical energy available for stimulating the nerve endings in the cochlea is, even for the everyday range of sounds, very small indeed. The tiny mechanical movements of the hair cells produced by such sounds release a store of energy many times that of the original sound.

Action Potential of Auditory Nerve Fibres.—The nature of the nerve impulses from the cochlea has been investigated at different levels in the afferent pathway. Galambos and Davis showed that in the second order cells of the cochlear nucleus there is a high sensitivity to tones of a definite frequency which is characteristic of each neurone. The selective response of these neurones is shown in Fig. 40, 11 where it can be seen that the frequency of discharge is raised with greater intensity of the tone, but the frequency for maximum sensitivity remains unaltered. Galambos has also shown that, in the cochlear nucleus, the response to one sound stimulus could be inhibited by the application of another sound stimulus. This type of mechanism would provide a means whereby the second-order neurones could be made highly selective in their response to particular tones.

Much interest therefore attaches to the response of the primary neurones, and this has recently been investigated in the guineapig by Tasaki. He gained access to the auditory nerve through a small hole drilled in the modiolus and was able to record the activity of single fibres in the nerve by pushing a glass capillary microelectrode of less than $\frac{1}{3}$ μ diameter at the tip into their myelin sheaths. Simultaneous records of the whole nerve response and of the cochlear microphonic, to sound stimulation, enabled him to clarify considerably the role of the primary neurones. The nerve fibres arising in the basal turn of the cochlea were found to respond to all audible frequencies, whereas those from the apical part responded to low frequencies only. The higher the frequency of the sound, the fewer were the fibres activated. Responses of several fibres, in terms of frequency and intensity of the sound stimulus, necessary to cause

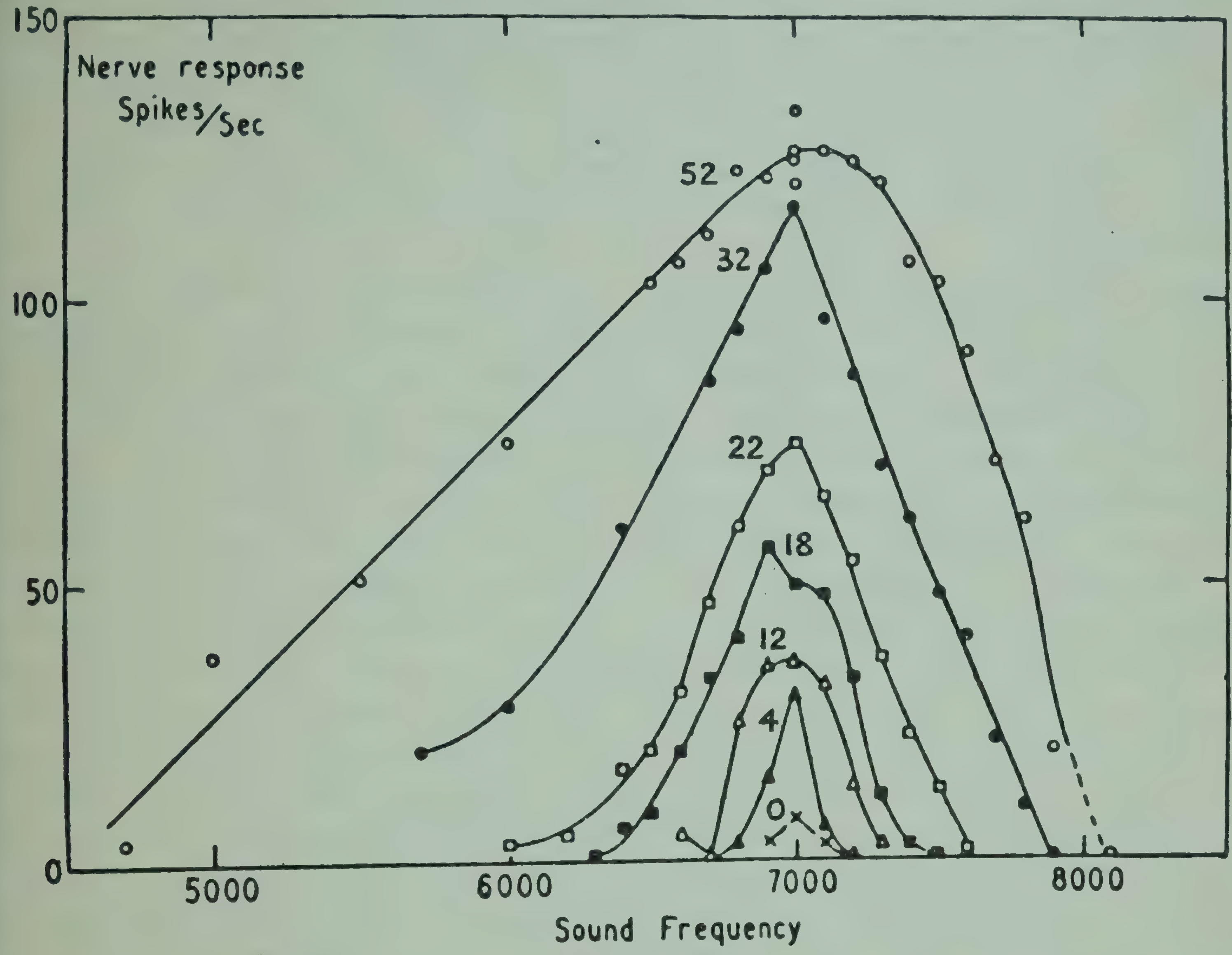


FIG. 40, 11

The relation between intensity of sound and the frequency of discharge in the second-order neurones of the dorsal cochlear nucleus. Each contour line shows the number of impulses recorded per second when a series of tones of the same intensity but of different frequencies, in the range 5000 to 8000 c/s, was applied to the cat's ear. The lowest curve has been marked 0 on the assumption that the intensity used was just threshold. The other curves were obtained with intensities 4, 12, 18, 22, 32 and 52 db above the threshold. The microelectrode in the experiment recorded here picked up the discharge from a neurone which showed a maximum response at 7000 c/s at all intensities of stimulation. *After* R. Galambos & H. Davis (1943), *J. Neurophysiol.* **6,** 39). Later experiments by Tasaki on the first-order neurones show a comparable cut off in the high frequency end of the response curve and a much more gradual fall towards the low frequency end.

activity in the fibre, were obtained. The curve shows a sharp cut-off at the high frequency end; that is to say as the frequency of the stimulating tone is increased a point is reached when the discharge of the neurones ceases comparatively suddenly. This steepness of cut-off at the high-frequency end of the response curves is of the same order as the curves of Galambos and Davis for the second-order neurones (Fig. 40, 11), but on the low frequency side there is no sharp limit to the response as in the second-order neurones. The travelling wave on the basilar membrane produced by a pure tone thus appears to die out at a definite point which depends on its frequency. The cochlear nerve might be thought of as signalling the place on the basilar membrane where the travelling wave suddenly dies out. On the basis of such information pitch could be discriminated.

Some auditory fibres discharge spontaneously, but this discharge, unlike that occurring in the second-order neurones, cannot be inhibited by sound stimuli. The occurrence of spontaneous discharges shows that the system is normally near the discharge point and that fluctuations in the level of activity of the excitatory processes are occurring. Thus in the presence of a constant acoustic stimulus some variation of response may occur; in general, however, the more intense the sound the higher is the frequency of discharge in the first-order neurones, the number of impulses being related approximately logarithmically to the sound pressure. With pure tones up to 1000 c/s the impulses in single nerve fibres tended to appear at the same frequency as the sound wave but the refractory period of the nerve fibre sets a limit to such synchronism at higher sound frequencies. Examination of the time relations between the cochlear microphonic and the occurrence of impulses in the auditory nerve fibres showed that, at the lower frequencies, nerve impulses were set up early in that part of the cochlear microphonic when current is flowing from the scala vestibuli to the scala tympani. This is further evidence for the belief that the microphonics are the direct electrical stimulus to the endings of the auditory nerve fibres.

Synopsis of Modern View of Action of Cochlea

It seems that the basilar membrane responds to acoustic stimulation with a pattern of vibration in which even pure tones excite a considerable length of the membrane; travelling waves occur which die out at various distances from the base, the higher frequencies being eliminated nearer the base, the lowest frequencies reaching the apex. In this way are produced the cochlear microphonics which, in general, correspond to the pattern of mechanical excitation of the membrane; the microphonics in turn stimulate the endings of the auditory nerve on the hair cells, setting up a pattern of activity in the fibres of the nerve. Higher intensities produce higher rates of discharge in the nerve fibres. The pattern of discharge is modified by inhibitory processes operating in the cochlear nucleus so that the second-order neurones show greater selectivity in their response to different frequencies of acoustic stimuli. It is also probable as suggested in the previous paragraph, that at low frequencies (but not at high frequencies) the sense of pitch is conveyed by the frequency of the nerve impulses.

Central Connexions of the Cochlear Nerve

On entering the medulla each fibre of the cochlear nerve divides into an ascending and a descending branch, the fibres from the apex of the cochlea having their bifurcations highest, while the lowest bifurcations are of nerves from the base of the cochlea. In this way a 'map' of the cochlea is formed in the brain stem. A similar map is found in the medial geniculate body and in the auditory radiations going from it to the cerebral cortex. The ascending branch goes to the ventral cochlear nucleus and the descending branch to the dorsal cochlear nucleus in the auditory tubercle. From

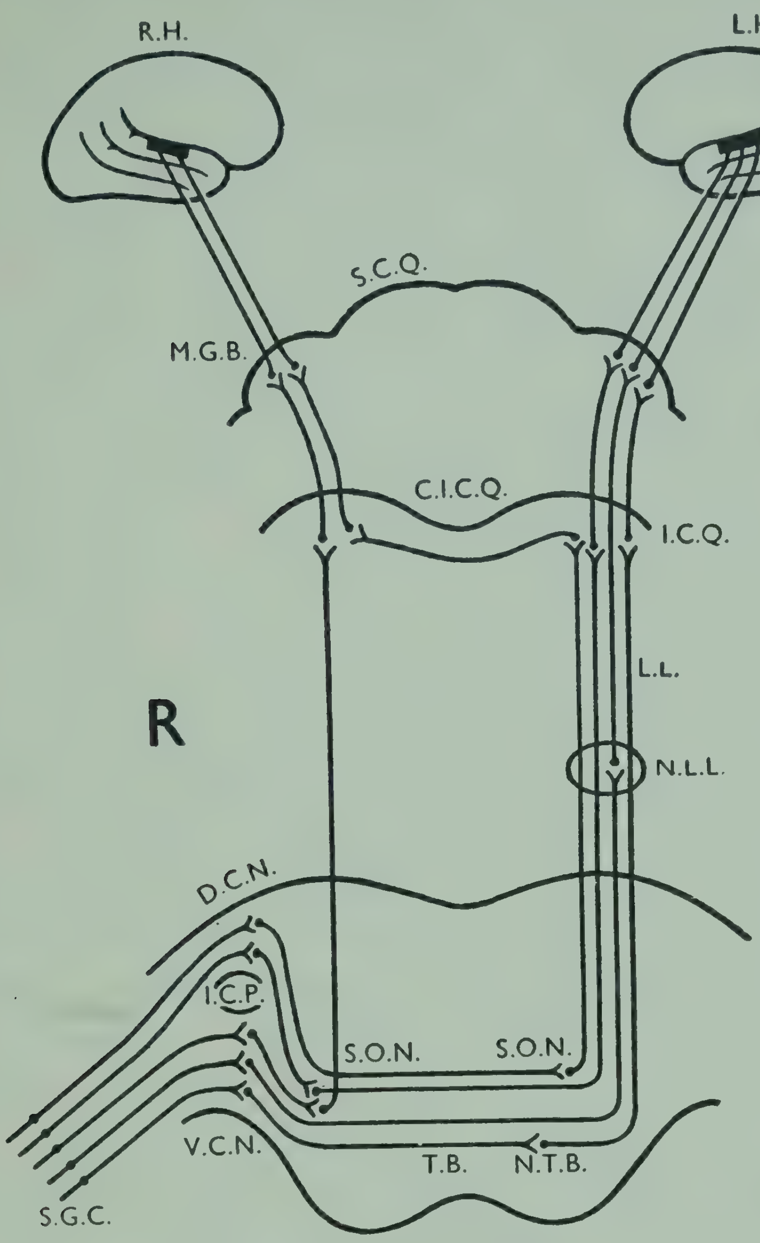


Fig. 40, 12

R.H. = Right hemisphere.
L.H. = Left hemisphere.
M.G.B. = Medial geniculate body.
S.C.Q. = Superior corpus quadrigeminum.
I.C.Q. = Inferior corpus quadrigeminum.
C.I.C.Q. = Commissure of the inferior corpus quadrigeminum.
L.L. = Lateral lemniscus.
N.L.L. = Nucleus of lateral lemniscus.
D.C.N. = Dorsal cochlear nucleus.
V.C.N. = Ventral cochlear nucleus.
I.C.P. = Inferior cerebellar peduncle.
S.G.C. = Spiral ganglion of the cochlea.
S.O.N. = Superior olivary nucleus.
T.B. = Trapezoid body.
N.T.B. = Nucleus of the trapezoid body.

(*Courtesy of G. W. Pearce.*)

the cochlear nuclei second-order neurones arise, most of which cross the midline and pass up in the lateral lemniscus to the inferior quadrigeminal body and the medial geniculate body. Nearly all these fibres are interrupted by a synapse in the trapezoid body or in the nucleus of the lateral lemniscus. From the medial geniculate body new fibres pass in the posterior limb of the internal capsule to the superior temporal gyrus. The number of synapses can be calculated from the time between the cochlear microphonic and the electrical changes due to nerve impulses in the various tracts. The nerve impulses started off by the cochlear microphonic are conducted rather slowly along the fine fibres to the cochlear nerve and so the latent period of 0·9 msec between the microphonic and the cochlear nerve impulse is accounted for. In the second-order neurones of the trapezoid body it is 2·2 msec and in the third-order neurones of the lateral lemniscus about 3·7 msec. The difference in latent period is accounted for by synaptic delay (0·8 msec) and conduction time (about 0·5 msec). The minimum latency in the cortex is about 8 msec. There is clinical evidence that there is a map of the basilar membrane in the cortex; investigation by electrical methods in the monkey shows that the anterior part of the auditory cortex receives impulses from the apical turn of the cochlea, while the posterior part is connected with the basal turn (Fig. 49, 15).

Hearing has been tested in dogs by the method of conditioned reflexes after removal of various parts of the auditory pathway. Removal of one cerebral cortex scarcely affected auditory acuity but removal of both hemispheres produced a loss of about 70 db. (see also p. 902).

The classical auditory pathway, as just described, from the ear to the cortex is not the only one since cats in which this pathway had been cut in the mid-brain could readily be aroused from sleep by loud sounds. The nerve impulses must be presumed to pass in a parallel pathway in the mid-brain reticular formation. There are also ascending pathways, as shown electrically, from the cochlea to the vermis of the cerebellum and from there to the cerebrum. In the opposite direction fibres descend from the olive to the cochlea and seem to end on the inner hair cells, but no function has yet been ascribed to them.

Localization of the Apparent Source of a Sound

The position of the source of a sound may be detected in either of two ways. If the head is kept still the observer can point to the source of the sound right or left of the saggital place. Alternatively the head may be moved from side to side until the sound seems to be directly ahead. The second is the more accurate method, and can be made even more accurate if the ears are extended by trumpets, as in the apparatus used to locate aeroplanes before radar was available. In practice the eyes as well as ears are used in localization. If the source cannot be seen, localization by hearing alone is sometimes quite difficult.

The ability to localize the source of a sound depends partly on the difference in loudness at the two ears and partly on the difference in time of arrival at the two ears. If the wavelength much exceeds the width of the head the difference in time of arrival at the two ears is the main factor (Fig. 40, 13, A) and the difference of intensity is scarcely detectable. When the frequency is such that the wavelength is nearly the same as the width of the head (about 1500 c/s) this method fails, as shown in Fig. 40, 13, B. Above 1500 c/s the head casts a 'sound shadow,' as these short waves cannot get round the head, just as the waves of the sea cannot get round a long breakwater; consequently if the source of a high-frequency note is to one side of the median plane there is a considerable difference of intensity at the two ears, which allows of localization.

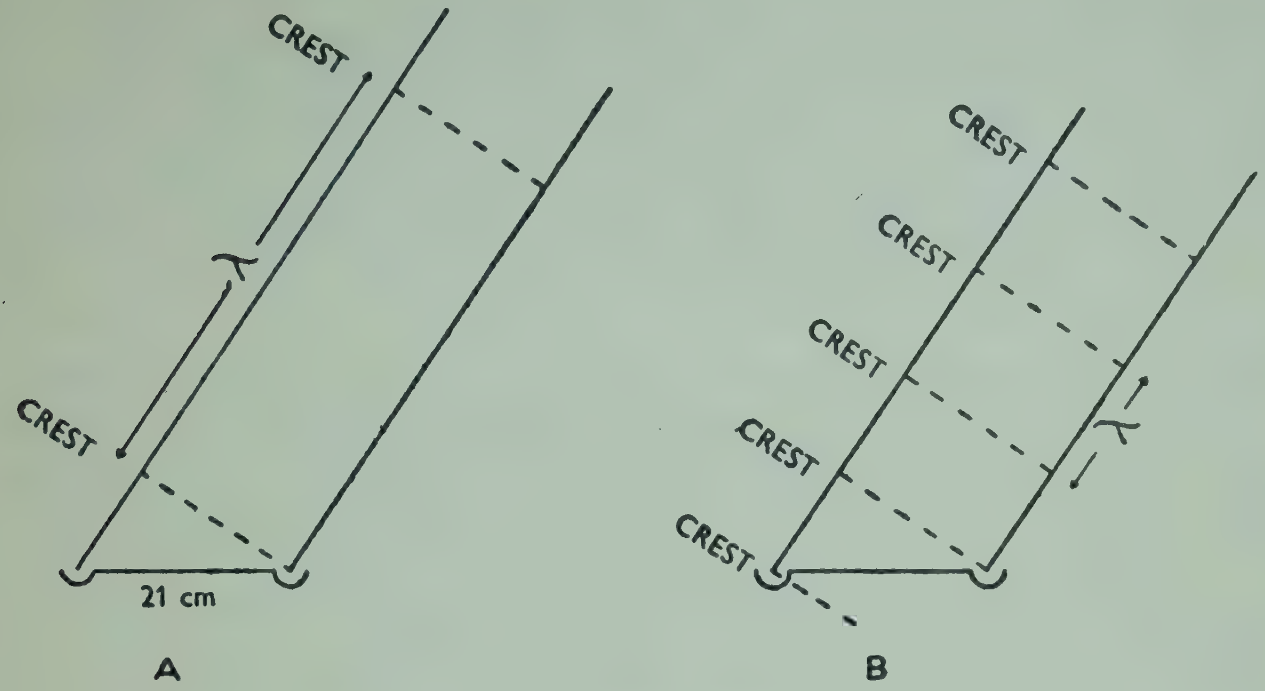


FIG. 40, 13

A. shows how a sound source is located by the crest of a wave (or some other distinctive part) reaching one ear before the other. The semicircles at the ends of the 21 cm horizontal line represent the ears. B. shows how this method breaks down when the wavelength is short compared with the distance between the ears, that is 21 cm. In this case one crest arrives at the left ear as the following crest reaches the right ear. The velocity of sound in air is approximately 330 metres per second at 0° C and varies directly as the square root of the absolute temperature of the air.

If the subject S, in Fig. 40, 14 is placed 3 metres from two loudspeakers, LS_1 and LS_2, which are producing the same sound with equal intensity and without phase difference the sound seems, as would be expected, to come from a point midway between LS_1 and LS_2. If the sound from LS_2 is delayed from 5 to 35 msec the sound appears to the subject to be coming only from LS_1, as if LS_2 were not present. If the delay is over 50 msec the sound from LS_2 appears as an echo of that from LS_1. This is of importance when loudspeakers are distributed in an auditorium to reinforce the voice of an actor speaking on the stage. Provided that the sound from the loudspeaker arrives at the listener in the auditorium 5 to 35 msec later than the sound travelling direct from the actor to the listener and that the sound from the loudspeaker does not exceed the power of the direct unamplified speech by 7 db, the illusion that the sound is coming from the actor is preserved.

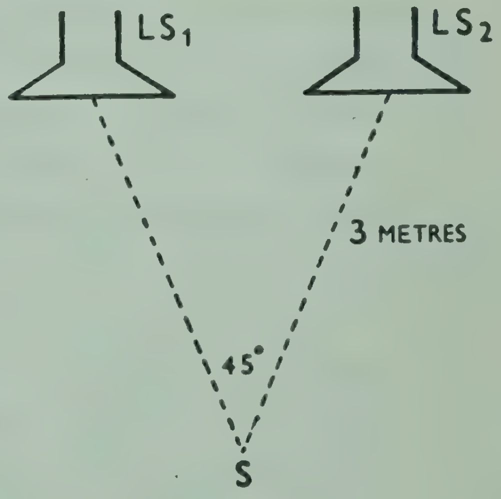


FIG. 40, 14

Diagram to illustrate echo and sound reinforcement.

Methods of measuring Acuity of Hearing

The traditional method of testing hearing by finding if the patient can understand spoken commands uttered at various distances and intensities has the advantage of requiring no apparatus but the great disadvantage that it cannot be accurately standardized. Hearing may also be tested by finding the maximum distance at which the tick of a watch can be heard but this cannot be standardized; a further objection is that the tick contains only a narrow range of frequencies, usually in the region of 2000 c/s.

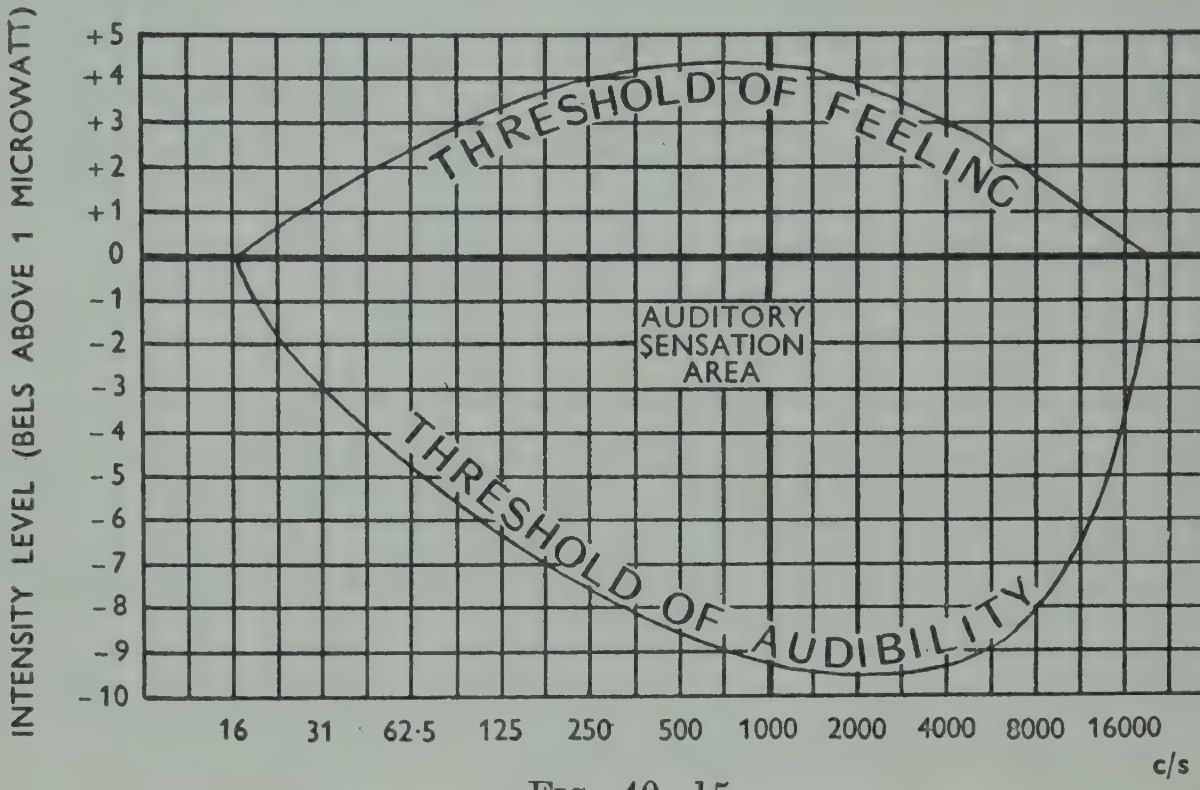


Fig. 40, 15

Diagram to show average threshold of audibility and average threshold of feeling (H. Fletcher (1929), "Speech and Hearing," Fig. 82, p. 157. London: Macmillan).

Quantitative results can be obtained with a *pure-tone audiometer* and even then only if the test is made in a sufficiently quiet room, since extraneous sounds cause masking (p. 728) which raises the threshold of hearing. The pure-tone audiometer consists of a calibrated valve oscillator producing pure sine waves which are fed to a telephone earpiece through an attenuator network to control the intensity of the sound. When the intensity of a given note is made very small it is inaudible but if the intensity is increased slowly then at a certain value the note is heard. This is the *threshold of audibility*. If the intensity is gradually increased the note becomes louder and louder. In the middle range of frequencies a very intense sound causes a tickling sensation which, as the intensity is increased, becomes actually painful. This is the *threshold of feeling*. In the lower frequencies the sensation is not usually so painful but intense sounds produce a fluttering sensation. Since intensities of this order can damage the organ of Corti, the threshold of feeling is the practical upper limit for sounds which are to be appreciated by the ear. The thresholds for any one person are scattered irregularly but the values lie near the smoothed curves shown in Fig. 40, 15. The area between the two curves is the *area of useful hearing*, or the *auditory sensation area*. In the standardization of apparatus it is now agreed that the standard threshold of hearing shall be taken as the intensity corresponding to 0·0002 dyn/cm^2 at 1000 c/s.

The threshold of audibility is lowest, that is the auditory acuity is greatest, in the region 1000 to 4000 c/s; at lower and higher frequencies the sound intensity must be greater to produce an auditory sensation. The lowest and highest audible frequencies thus depend on the intensity of the tone used, but the average limits are 20 and 20,000 c/s. In old people the threshold in the higher frequencies is raised so that in practice their upper limit may be much lower, say 5000 c/s (see Fig. 40, 17. 1).

Some idea of the great sensitivity of the ear is given by the following figures. The ear is most sensitive to sounds of frequency of about 3500 c/s at which the threshold in a quiet room is approximately 0·0001 dyn/cm.^{2} This is produced by an oscillatory movement of the air particles from their mean position of $1{\cdot}25 \times 10^{-10}$ cm, which is about one-hundredth of the diameter of a molecule of hydrogen.

By varying the intensity of the note produced by a pure-tone audiometer in a rhythmic fashion the minimum perceptible difference in intensity can be measured. Except near the threshold of audibility it is of the order of 1 db. By varying the frequency of a note rhythmically the minimum fractional difference in frequency which is perceptible can be found; in the range 500 to 4000 c/s it is 0·3 per cent., but outside this range it is greater. From this it is calculated that there are upwards of 2000 gradations in pitch perceptible to the human ear. A tone must persist for 10 to 15 msec before a pitch can be assigned to it but most people require the tone, especially if it is of low frequency, to last upwards of 100 msec before it supplies an experience of pitch. It is impossible to say where pitch is recognized as removal of the auditory cortex in the dog or monkey has little effect in discrimination of tones.

Frequency distortion is a common fault of all communication systems. It means that all the frequencies present in the original sound are not amplified equally. In the post office telephone the higher frequencies are poorly reproduced and the quality of the speech is poor, so that the intelligibility is not so high as it is in technically good reception of transmissions from broadcasting stations, in which the reproduction of the higher frequencies is good.

The method of measuring intelligibility, first used by telephone engineers, is also useful for investigating deafness and the efficiency of deaf aids. A list of disconnected words containing a wide variety of speech sounds but without context is read out to a listener who writes down what he hears. The script is then checked with the original list and the percentage of sounds heard correctly—spelling errors are of course neglected—is usually referred to as the 'intelligibility.'

When this test is made in the simplest possible conditions—that is, with the reader speaking directly to the listener in the same room—the intelligibility is high, but the vowel *e* as in 'ten' and the consonants *th*, *f* and *v* often give rise to errors. If now the listener hears the voice of the reader through a loudspeaker system which transmits all the speech frequencies up to 1000 c/s but cuts out all the higher frequencies the intelligibility will be only 40 per cent., although almost all the energy (85 per cent.) of the original speech sounds is present. Sounds like *s*, *th* and *f* are particularly difficult to distinguish. This is an artificial high-frequency deafness. If, on the other hand, the loudspeaker transmits only the speech frequencies above 1000 c/s the intelligibility is now

86 per cent., although only a small fraction of the original speech energy (17 per cent.) is transmitted. In spite of this severe frequency distortion ordinary speech with its context is easily understood although it does not, of course, sound natural. Music heard through this apparatus has the fundamental frequencies eliminated but the pitch of the notes is not altered and the various instruments are easily recognized. These findings emphasize once again the great importance of the higher frequencies.

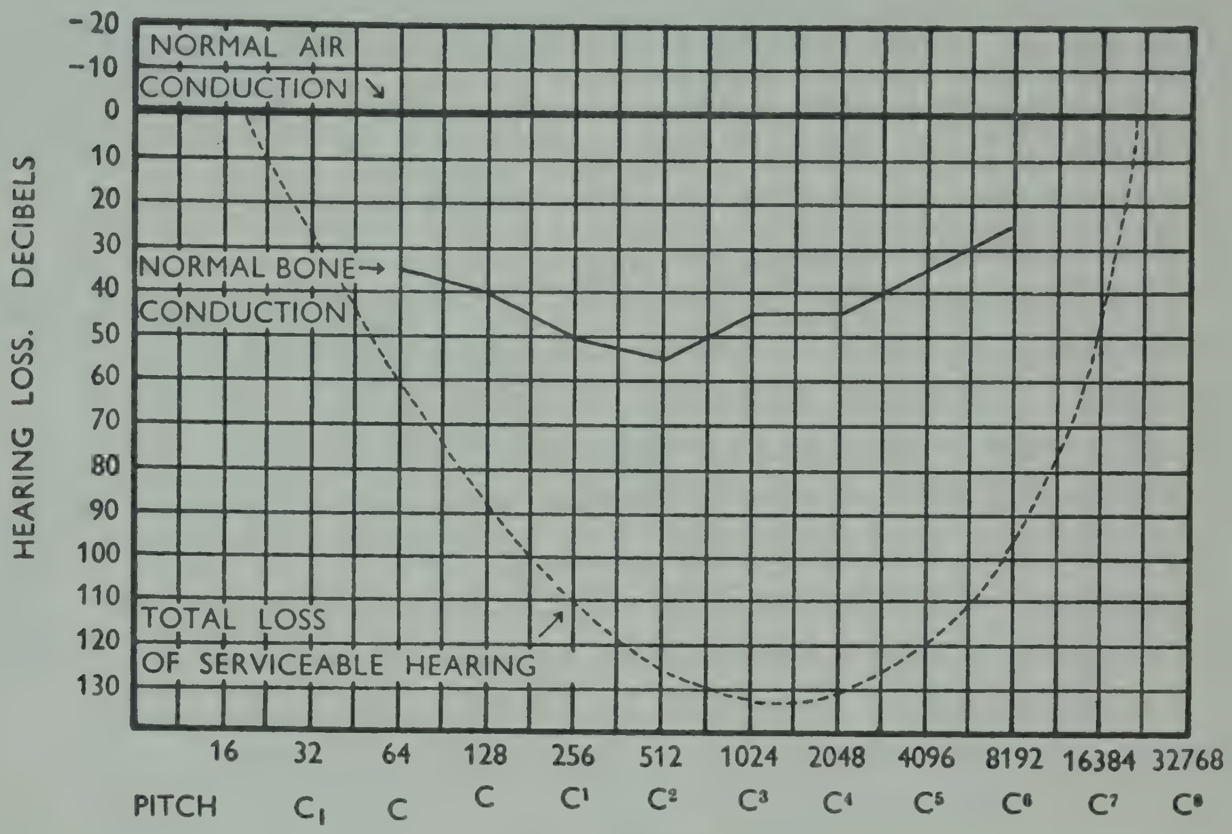


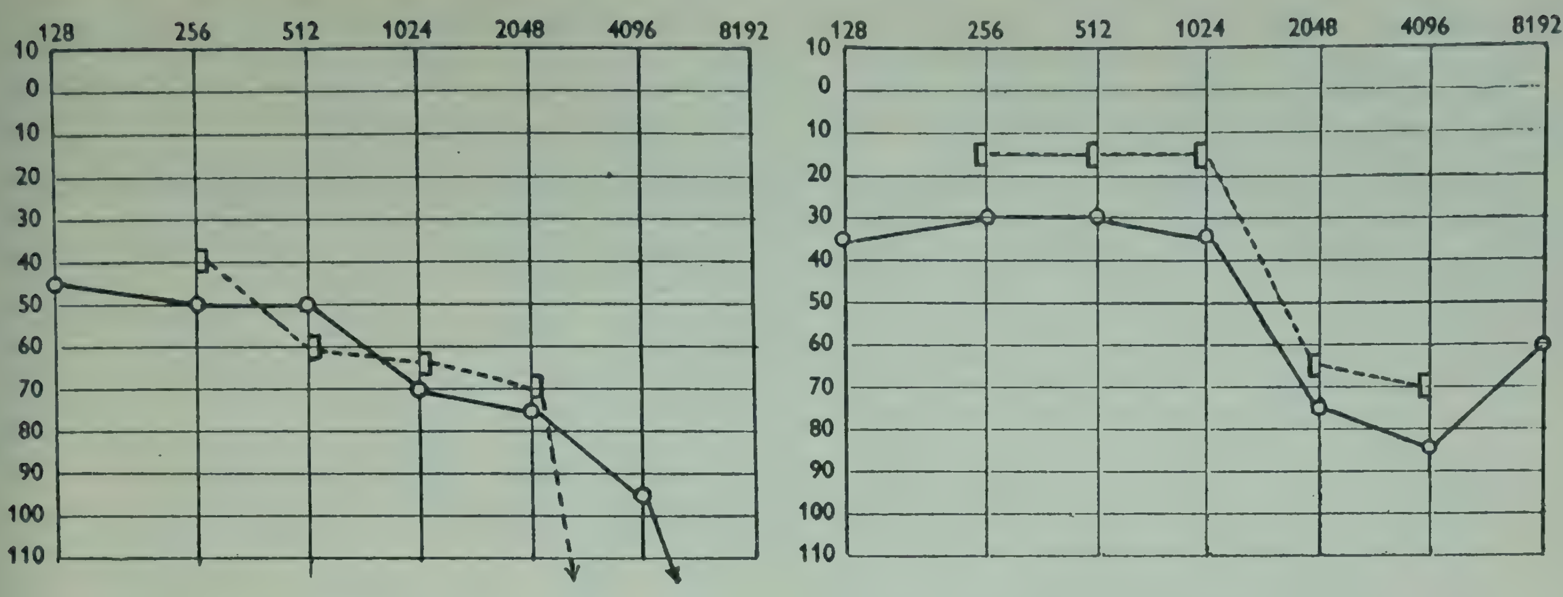
FIG. 40, 16

Chart for recording hearing loss. Audiogram chart (*Children with Defective Hearing* (1938), p. 41. *H.M.S.O.*). The thick line marked 'normal air conduction' is the threshold of hearing while the dotted line marked total loss of serviceable hearing is the 'threshold of feeling'.

(*By kind permission of Controller H.M. Stationery Office.*)

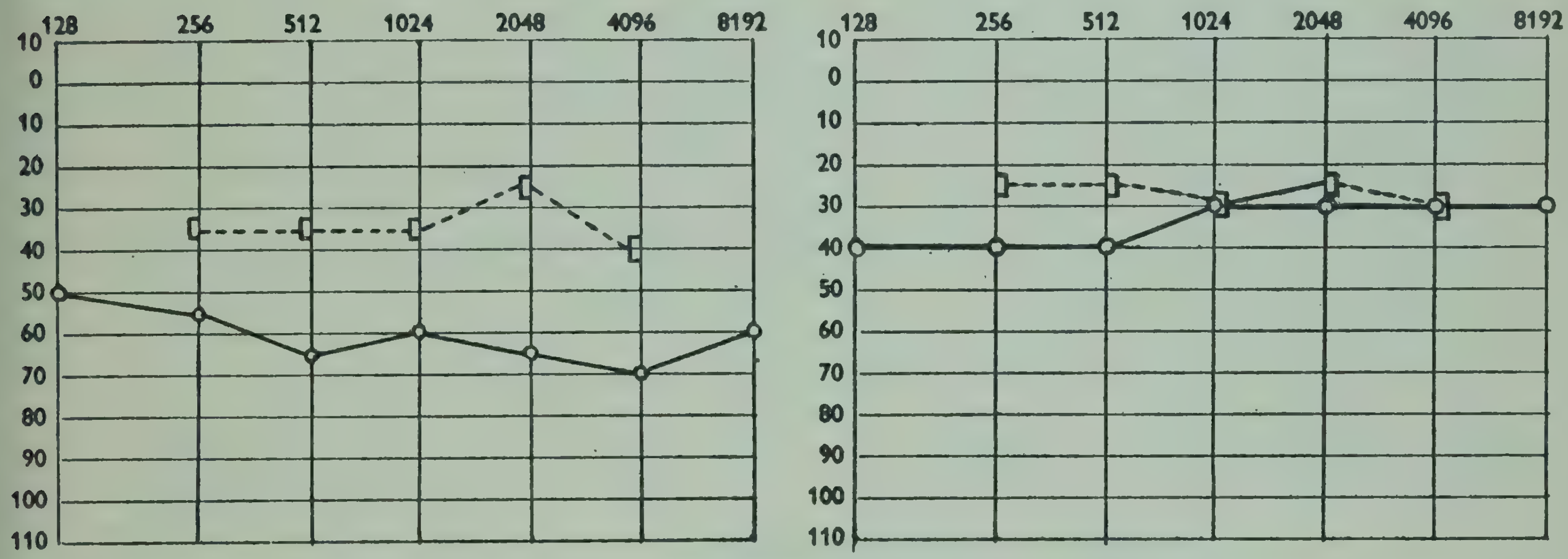
The pure tone audiometer can be used to investigate the hearing of deaf persons. Useful diagnostic information is obtained by measuring the amount by which the threshold is raised over a wide range of frequencies. Audiometers intended for clinical use are calibrated so that the zero mark on the intensity control at each frequency corresponds to the average threshold found in young people with normal hearing. This statistical concept of normal hearing is shown on the graph paper, used to record the results, as a straight line marked 0 db hearing loss (Fig. 40, 16). The threshold of feeling now appears on the lower part of the graph paper as "total loss of serviceable hearing." If a person is partially deaf to a particular tone its intensity must be increased x db above the average threshold before it is heard; this is plotted as a point *below* the normal threshold in Fig. 40, 16, and can be regarded as a *hearing loss* of x db at the frequency tested. If a number of tones is used the results fall into a curve called the audiogram. A few characteristic audiograms are shown in Fig. 40, 17.

Important prognostic information can be gained from the audiogram, since it has been found that there is a definite relationship between the degree of deafness in the important speech frequencies, say 500 to 2500 c/s, and the quality of the speech. Deaf children learn to speak normally provided the loss is less than 30 db; a greater degree of deafness prevents



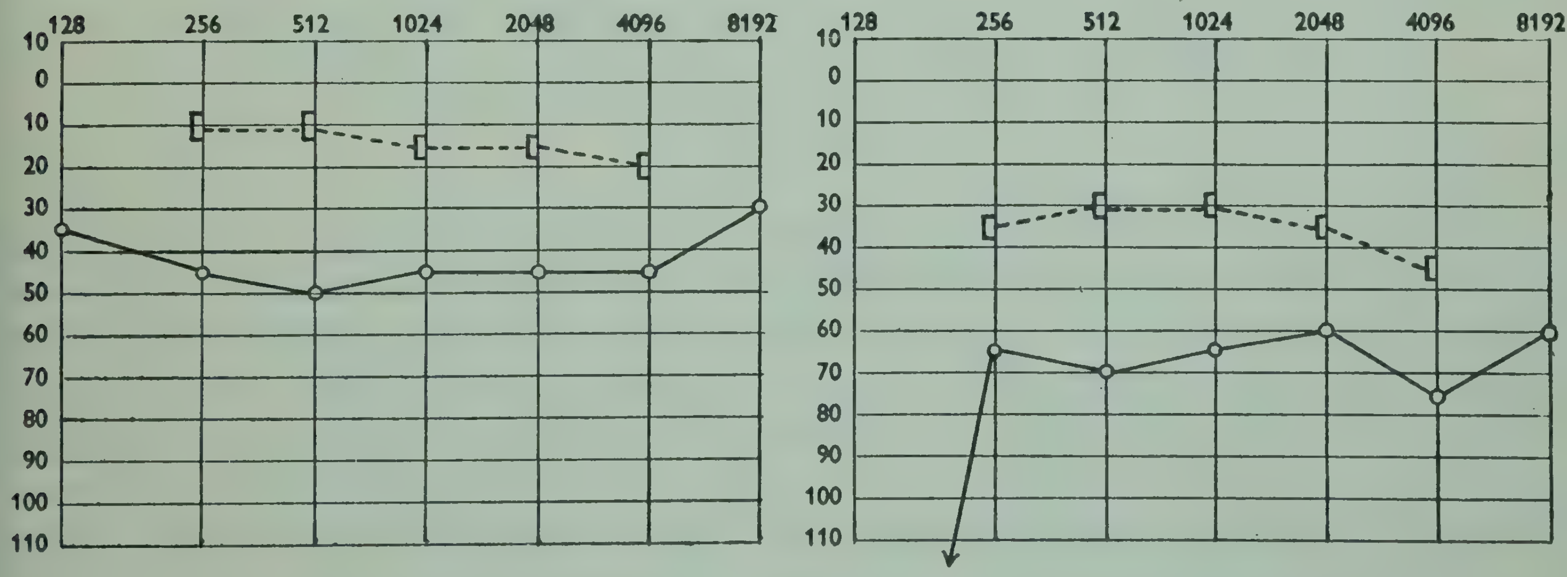
1. Well-marked presbycusis.

2. Traumatic deafness in a gun-layer.



3. Large central perforation in right tympanic membrane after measles.

4. Same patient. Audiogram of left ear for comparison.



5. Otosclerosis.

6. Tympanic membrane, malleus and incus removed at radical mastoid operation.

]- - - - -] Bone conduction.   o———o——— Air conduction.

Fig. 40, 17

Audiograms from five different patients which illustrate conditions mentioned in the text. Air conduction was tested by an ear-phone placed over the external auditory meatus. Bone conduction was tested by a vibrator placed on the mastoid process. O on the ordinate indicates the threshold of hearing in young healthy people. (*By courtesy of M. J. Gibson.*)

an infant from learning to speak naturally but may permit him to acquire speech with careful training. In the presence of a loss of 90 db, however, even acquired speech is poor. Good hearing in the lower frequencies (up to 500 c/s, say) does not compensate for deafness in the high frequencies, since it is the latter which allow us to distinguish the different speech sounds.

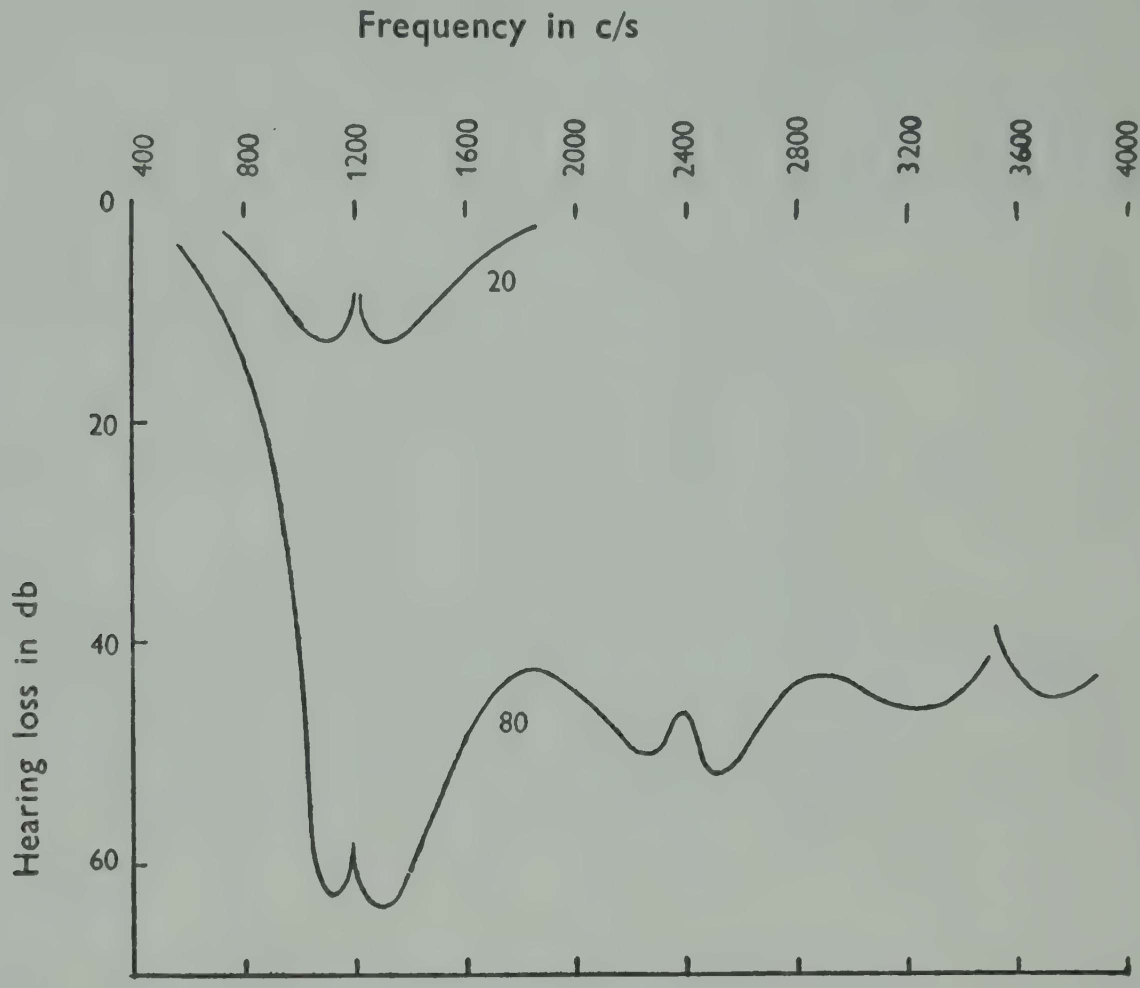


FIG. 40, 18

Audiograms of the hearing loss (monaural masking) at different frequencies of pure tones in the presence of a tone of 1200 c/s at intensities of 20 and 80 db above its threshold of audibility. 0 on the ordinate represents the threshold of audibility. Note that the frequency scale is linear and not logarithmic as in Fig. 40, 16. (*After* H. Fletcher (1929), *Speech and Hearing*. London: Macmillan.)

Deafness due to middle-ear disease is so common that it is important to be able to test large numbers quickly. The pure-tone audiometer is not suitable because it is a rather uninteresting artificial test requiring a special room and considerable time; it can be applied only to one person at a time. The gramophone audiometer provides a somewhat empirical but yet practical test which can quickly pick out persons with defective hearing. Specially made records are played with an electrical pick-up which is connected to twenty or more single earphones to allow a whole class of children to be tested simultaneously. The children apply the earpiece to one ear at a time and hear a voice calling out numbers which they copy down. As the record proceeds the voice becomes fainter and fainter. When the speech is too faint to be understood the children either make mistakes or stop writing. The performance of the normal child being known, it is easy to pick out even slight hearing defects.

Masking.—Whenever there is a noise (as in a railway train) conversation becomes difficult and the speech power must be raised. The increase in the loudness of the voice is often revealed if the noise suddenly ceases.

The nature of this masking effect has been investigated by means of the pure-tone audiometer. First, in an absolutely quiet room the threshold of audibility is measured, then a masking tone is continuously sounded and the threshold is measured again; the results are then plotted in an audiogram chart (Fig. 40, 18). If the intensity of the masking tone, in this case 1200 c/s, is not very great (20 db above its threshold), then, as might be expected, there is some difficulty in hearing notes of neighbouring frequency, but if the masking tone is much louder (80 db above its threshold) there is severe masking of the higher frequencies as well as of neighbouring frequencies. (The kinks in the curves of Fig. 40, 18 at 1200, 2400 and 3600 c/s are due to beats occurring between the masking tone or its harmonics and the masked tone.) Masking, since it affects the high frequencies, tends to impair intelligibility. In an electrical deaf aid the amplification of the lower frequencies should not be so great as to mask out the higher frequencies which allow us to distinguish between the various speech sounds (Fig. 40, 19).

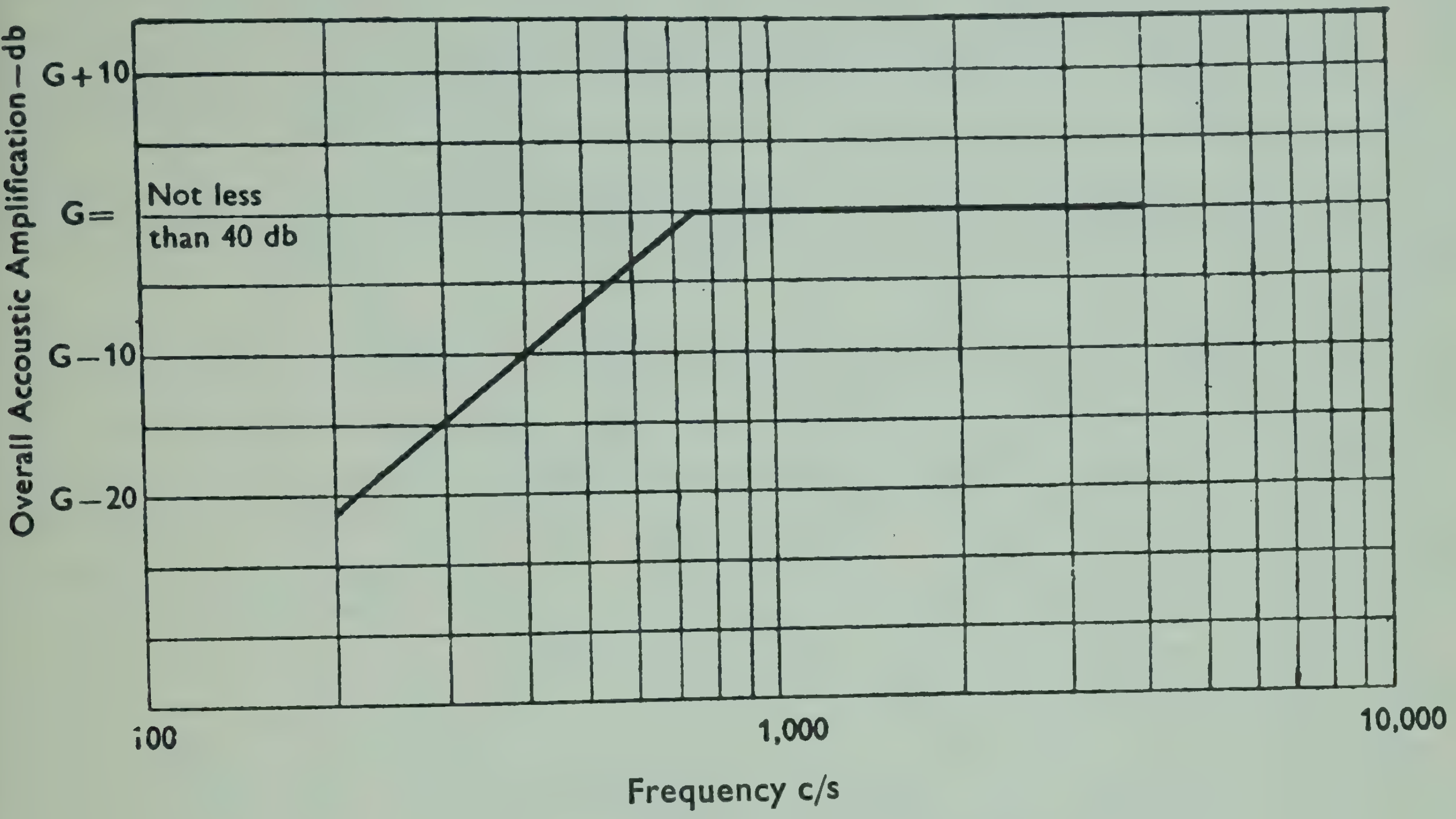


Fig. 40, 19

Amplification curve of deaf aid for general use recommended by the Medical Research Council. Note the lower amplification in the bass which reduces masking of the higher frequencies and is likely to give maximum intelligibility (*Hearing Aids and Audiometers* (1947), Spec. Rep. Series, 261, London, H.M.S.O.).

(*By kind permission of Controller H.M. Stationery Office.*)

In the experiments just described both the masking and the masked tone arrive at the same ear—this is described as monaural masking. In testing binaural masking the continuously sounded masking tone is applied to one ear while the threshold of the other ear is measured at various frequencies. To achieve the same masking effect under these conditions the masking tone may have to be 50 db louder than it was in the monaural experiments. There is, therefore, no alteration of the threshold of one ear unless the masking tone applied to the other ear is more than 50 db above its threshold. The masking is in fact due to the transmission of the masking tone by the skull to the opposite cochlea. This explains how a telephone conversation can be carried on in a noisy room and how faint heart sounds can be heard with a monaural stethoscope in spite of the noise in the ward. Since the interference of room noise is not due to noise entering the free ear, but to noise entering the ear to which the earpiece is applied, care must be taken to fit this firmly and closely to the ear.

Rinne's and Weber's Tests.—Sound vibrations can be conducted to the inner ear through the bones of the skull, but normally air conduction is much more effective. This can be shown by applying a vibrating tuning-fork (256 c/s) firmly to the mastoid process just behind the ear. When the sound dies away the prongs of the fork are brought near the external auditory meatus and the sound is heard once more, provided that the meatus is not blocked and that the ear drum and ossicles are intact, that is, Rinne's test is positive. That this finding probably depends on the masking effect of room noise may be shown in the following way. When the sound is no longer heard from the vibrating fork on the mastoid process, plugging the external ear with the finger allows the sound to be heard once more.

Weber's test is carried out by placing a vibrating 256 c/s fork on the mid-sagittal line of the head (the vertex). If one ear is plugged by the finger to exclude room noise the sound produced by the fork appears to be louder in that ear. Thus in unilateral disease affecting the conducting system (middle ear disease) Weber's test is referred to the diseased ear. When the organ of Corti or the cochlear nerve is damaged unilaterally Weber's test is referred to the normal ear.

Rinne's and Weber's tests are used in clinical research to distinguish middle ear (conduction) deafness from inner-ear (perceptive) deafness. In conduction deafness Rinne's test is usually negative and Weber's test indicates the diseased ear. In perceptive deafness Rinne's test is positive and Weber's test indicates the normal ear.

Noise

Noise is not to be defined in physical terms but rather in psychological terms such as a sound which is disturbing or annoying. The loudness of sounds can be measured either subjectively or objectively. In the subjective method the loudness of a sound is compared with the loudness of a note of 1000 c/s which can be adjusted until the listener judges that it has the same loudness as the sound being investigated. When this equality has been achieved the intensity of the 1000 c/s note can be read from the instrument in db above an arbitrary zero which has been internationally agreed as 0·0002 dyn/cm.^{2} If the reading is n db the sound is said to have an intensity of n phons. It is more usual to employ objective sound level meters which have a microphone and a calibrated amplifier with a number of alternative frequency response characteristics based on those of the ear at different sound levels. Such instruments give a reading of the sound level in db above the same reference level of 0·0002 dyn/cm.^{2} This objective method does not necessarily give a true index of loudness.

Auditory and Non-Auditory Effects of Noise

The effect of noise on man varies according to the intensity of the noise and on the circumstances of the exposure to it. At low levels of noise speech sounds may be masked, distraction and annoyance may be

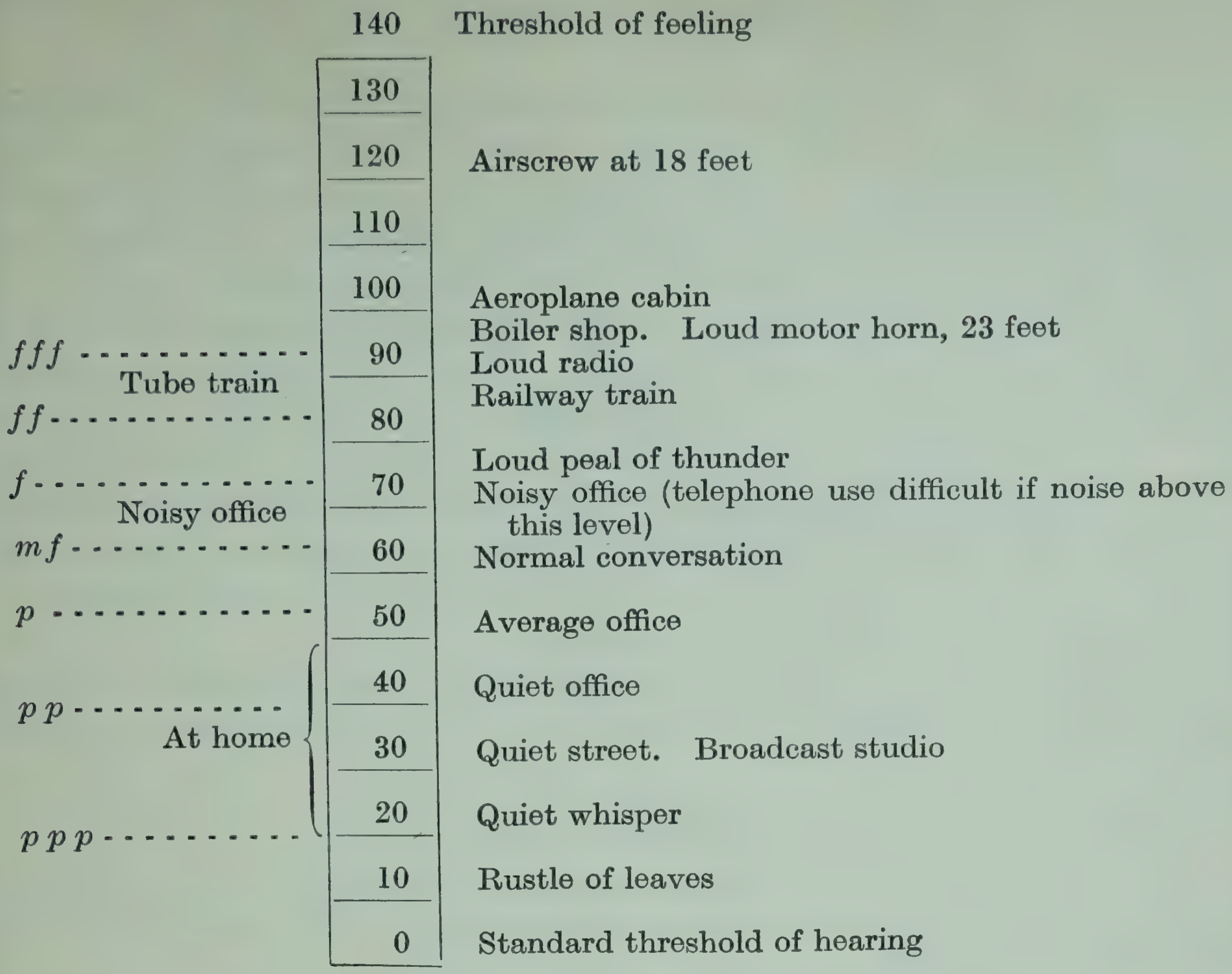

140 Threshold of feeling
130
120 Airscrew at 18 feet
110
100 Aeroplane cabin
Boiler shop. Loud motor horn, 23 feet
90 Loud radio
Railway train
80
Loud peal of thunder
70 Noisy office (telephone use difficult if noise above this level)
60 Normal conversation
50 Average office
40 Quiet office
30 Quiet street. Broadcast studio
20 Quiet whisper
10 Rustle of leaves
0 Standard threshold of hearing
fff
Tube train
ff
f
Noisy office
mf
p
pp
At home
ppp

Fig. 40, 20

A noise ladder scaled in phons. On the left are given the values of the *p* (piano) and *f* (forte) markings as used by Leopold Stokowski.

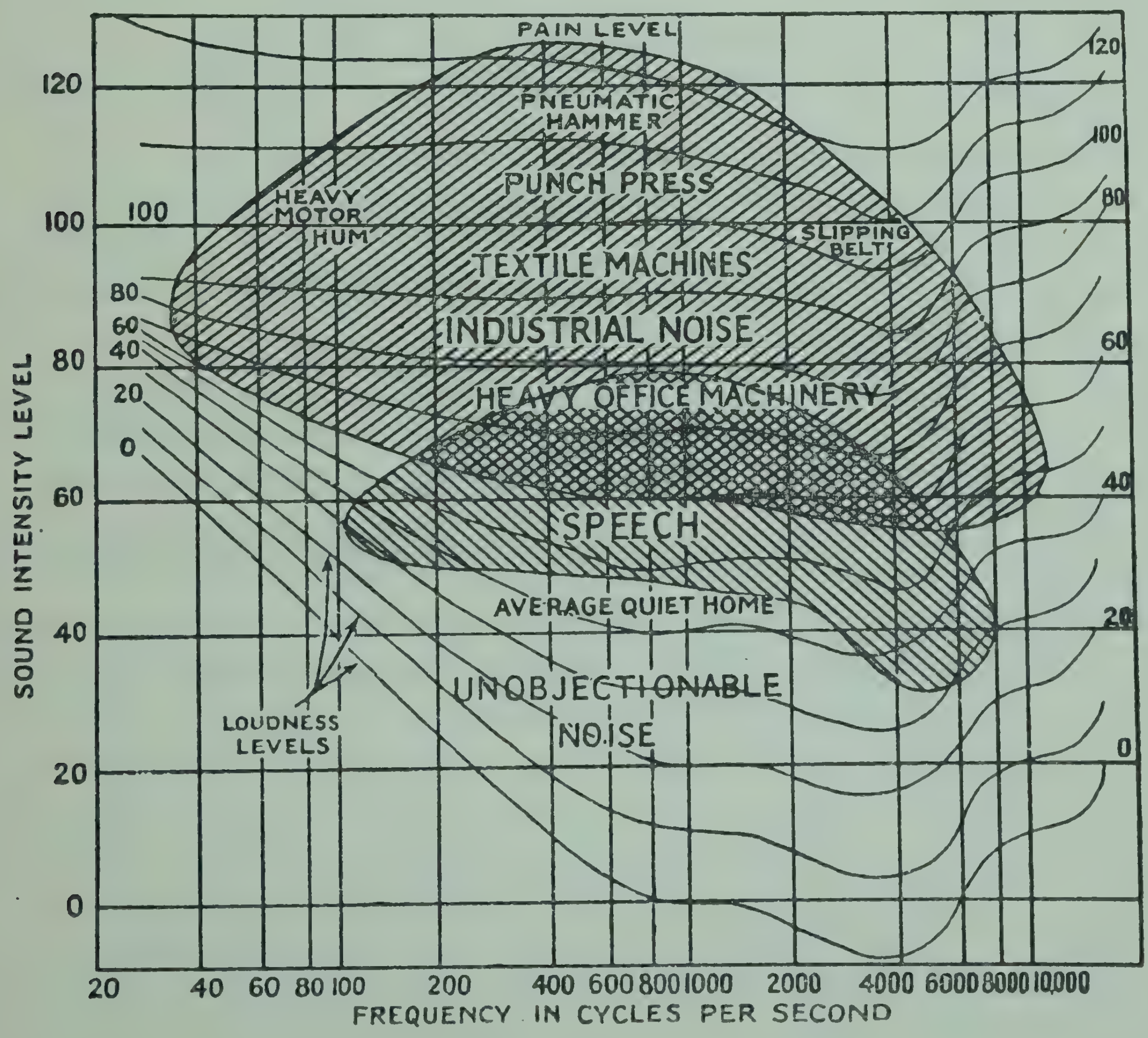


Fig. 40, 21

The distribution of various noises throughout the auditory sensation area. The curves marked 'loudness levels' are equal loudness levels and are marked in phons. The line marked 0 is the threshold of hearing. The ordinate gives the pressure in db above 0·0002 dyn/cm^2. (*After* H. C. Hardy (1947), *Industry and Power*, **52**, 77.) The equal loudness curves are from H. Fletcher & W. A. Munson (1933), *J. Acoust. Soc. Amer.* **6** 82–108.

experienced, but it is not easy to demonstrate any reduction in the efficiency of the performance of a wide variety of tasks. Broadbent, however, has shown that, in the presence of fairly loud noise maintained for some time, deterioration in the performance of complex tasks occurs in certain people.

It has been clearly demonstrated that persons exposed daily to noise, as in certain noisy occupations, without ear protection show hearing losses at first temporary but later permanent. The extent and rate of development of the deafness varies with different levels and types of noise and depend on individual susceptibility. This effect is additional to the gradual loss of acuity to high tones (presbycusis) which occurs with advancing age. When the noise is extremely intense damage to hearing quickly ensues and other effects, such as giddiness, due to stimulation of non-auditory parts of the labyrinth, are seen.

Noise, therefore, constitutes a danger to hearing which should be guarded against by reduction at the source or by the use of ear protection.

REFERENCES

Bast, T. H. & Anson, B. J. (1949). *The Temporal Bone and the Ear.* Oxford: Blackwell.

Broadbent, D. E. (1953). Noise, paced performance and vigilance tasks. *Brit. J. Psychol.* (Gen. Sect.), **44,** 295–303.

Davis, H. (1953). The inner ear. *Advanc. Sci., Lond.* **9,** 420–425.

Dix, M. R., Hallpike, C. S. & Hood, J. D. (1949). Nerve deafness, its clinical criteria, old and new. *J. Laryng. Otol.* **63,** 685–698.

Fletcher, H. (1929). *Speech and Hearing.* London: Macmillan. Also 2nd Ed. (1953). London: Macmillan.

Galambos, R. (1954). Auditory pathways. *Physiol. Rev.* **34,** 497–528.

Potter, R. K., Kopp, G. A. & Green, H. C. (1947). *Visible Speech.* New York: D. Van Nostrand Co.

Russell, G. O. (1928). *The Vowel.* Ohio State University Press.

Stephens, S. S. & Davis, H. (1938). *Hearing: its Psychology and Physiology.* New York: Wiley.

Stevens, S. S., Ed. (1951). *Handbook of Experimental Psychology.* London: Chapman & Hall. Chap. 25–28.

Symposium on Neuro-otology (1956). *Brit. med. Bull.* **12,** 91–157.

Tasaki, I. & Davis, H. (1955). Electric responses of individual nerve elements in cochlear nucleus to sound stimulation (guineapig). *J. Neurophysiol.* **18,** 151–158.

Tasaki, I. (1954). Nerve impulses in individual auditory nerve fibres of guineapig. *J. Neurophysiol.* **16,** 97–122.

Watson, L. E. & Folan, T. (1949). *Hearing Tests and Hearing Instruments.* Baltimore: Williams and Wilkins.

Wever, E. G. & Laurence, M. (1954). *Physiological Acoustics.* Princeton University Press.

Wever, E. G. (1949). *Theory of Hearing.* London: Chapman & Hall.

CHAPTER 41

NEURONE AND SYNAPSE

NEURONE

THE unit of the nervous system is the nerve cell, or neurone, which is specially adapted for the carrying of nerve impulses. These impulses pass on from one neurone to another at junctions called *synapses*. The process may be likened to a relay race in which an impulse, brought to a synapse in the axon of one neurone, is relayed by a second neurone to the next synapse where the impulse may be relayed once more. Since the nervous system is essentially an enormous number of interconnected neurones supported in a framework of non-neural tissue it is necessary to study first the characteristics of the two elementary components, the neurone and the synapse.

For purposes of description a motor cell of the anterior horn of the grey matter of the spinal cord (Fig. 41, 6) may be taken as a typical nerve cell as found in vertebrates. The cell body contains a nucleus, usually with a well-marked nucleolus, surrounded by cytoplasm containing granules (Nissl's granules) which stain readily with basic dyes such as methylene blue. The cytoplasm is prolonged into numerous short branching *dendrites* (or dendrons) and a single *axon*, or axis cylinder, which may be very long. The axon leaves the cell body at the axon hillock where there are few Nissl's granules. Fixed and stained nerve cells and axons appear to have neurofibrils running through them but these may be artifacts produced by fixation. Microdissection of freshly excised frog nerves shows that the axons have the consistency of a soft jelly, probably consisting of oriented protein chains in a gel state surrounded by a cell membrane. About 100 μ after it has left the cell body, but while it is still within the spinal cord, the axon acquires a coat, the *myelin* or *medullary sheath*, and is then called a myelinated or medullated nerve fibre. The myelin sheath, which varies in thickness in different nerve fibres, consists of regular, concentrically arranged lamellæ with a period of approximately 180 Å of oriented lecithin molecules with some protein and water. Lecithin, a phospholipid, stains black with osmic acid (see also p. 823). After emerging from the central nervous system the fibre receives another covering, the *neurilemmal sheath*; this is thin and structureless and often continues right up to the muscle cells supplied by the axon. The myelin sheath, on the other hand, is broken up into short lengths or segments by constrictions of the neurilemma, called *nodes of Ranvier*, where, for a distance of 0·25 to 1 μ the axon is covered by neurilemma only (Fig. 41, 1B). In the node the diameter of the axon is slightly diminished. The distance between the nodes of Ranvier in a nerve fibre of 10 μ diameter is about 0·7 mm; the internodal distance is greater in larger fibres. The axon of an anterior horn cell has numerous branches which always arise at these nodes. A very thin layer of protoplasm, the *cell of Schwann*, covers each segment of myelin sheath; the oval nucleus of this cell is placed under the neurilemma about half-way between the nodes. As the fibre emerges from the central nervous system it acquires a further coating of very fine fibrous tissue, the endoneurium. There is a steady proximo-distal movement of fluid in the endoneurial spaces of peripheral nerves at the rate of a few mm per hour. The origin and fate of this fluid are, however, not known.

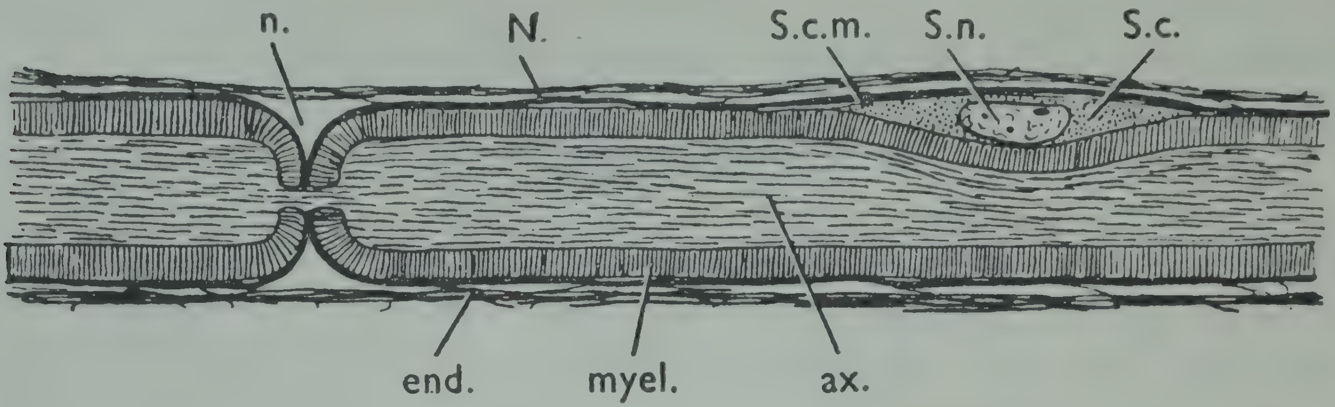


FIG. 41, 1 A

Diagram of part of a normal vertebrate nerve fibre. The Schwann nucleus lies nearer to the node than it would usually do, but otherwise the proportions are correct. ax., axon; end., endoneurium; myel., myelin sheath; n., node of Ranvier; S.c., Schwann cytoplasm; S.c.m., Schwann cell membrane; S.n., Schwann nucleus; N., neurilemma (W. Holmes & J. Z. Young (1942), *J. Anat.* **77**, 67).

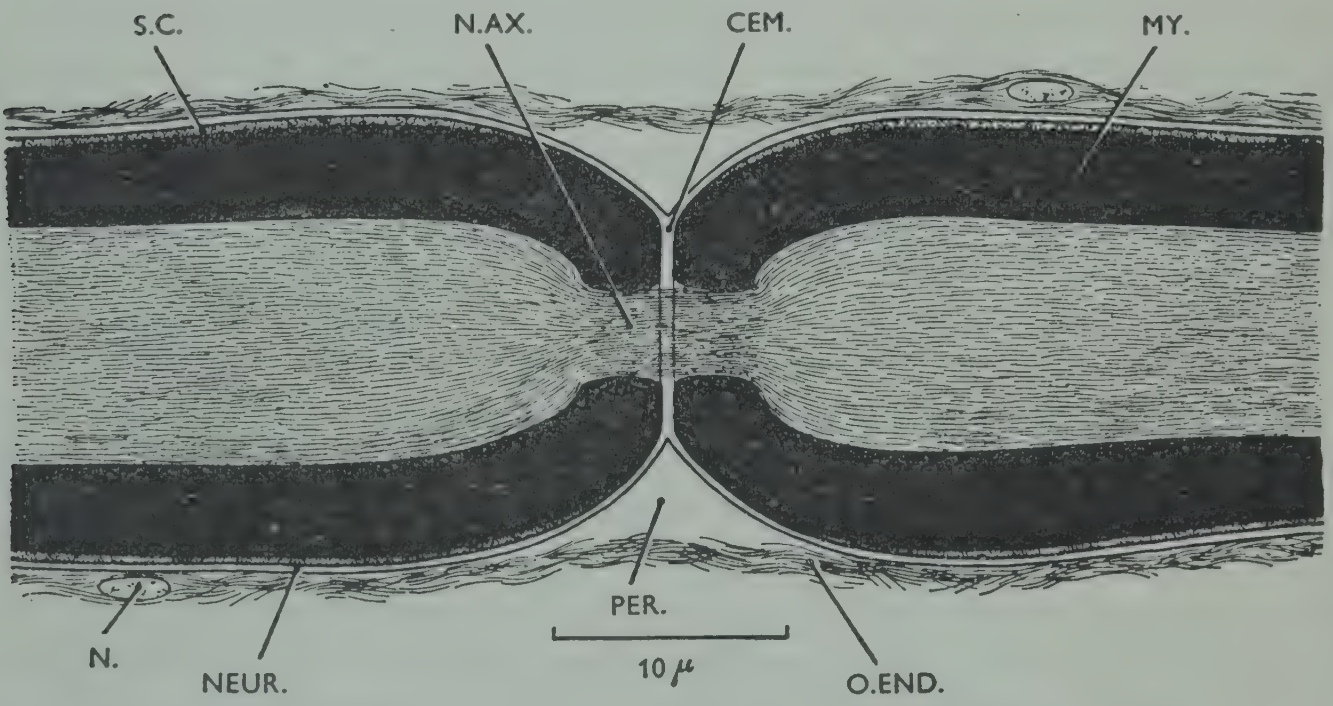


FIG. 41, 1 B

A diagram on a larger scale to give more detailed information about the nodal region. The reconstruction is based on a tracing from a photograph of a fibre fixed in osmium tetroxide, but details are added from preparations made by other methods. Any actual fibre would show greater irregularities in the outline of the myelin. *cem.*, cementing disc; *my.*, myelin; *n.*, nucleus of outer endoneurium; *n.ax.*, nodal axon; *neur.*, neurilemma; *o.end.*, outer endoneurium; *per.*, perinodal space; *s.c.*, Schwann cell protoplasm. (A. Hess & J. Z. Young (1952), *Proc. Roy. Soc.* B **140**, 306).

The description just given of a motor fibre applies also to a sensory fibre, as both have exactly similar histological appearances. In the case of the sensory fibre, however, the cell body does not lie within the spinal cord but in a swelling on the posterior root of the spinal nerve, the *spinal ganglion*. This contains unipolar nerve cells (Fig. 45, 3), each having a very short single process which divides into a central fibre going into the spinal cord and a peripheral fibre carrying impulses from the periphery (skin, joints, muscles and viscera).

In the peripheral nervous system of vertebrates there are also numerous very thin fibres which, since no myelin sheath could be demonstrated by ordinary microscopic methods, have been called non-myelinated fibres. However, we now know that they have a very thin layer of oriented fatty molecules, visible only with the high magnification of the electron microscope, forming the excitable membrane around the axon. In invertebrates myelinated nerve fibres have rarely been evolved but a few species have

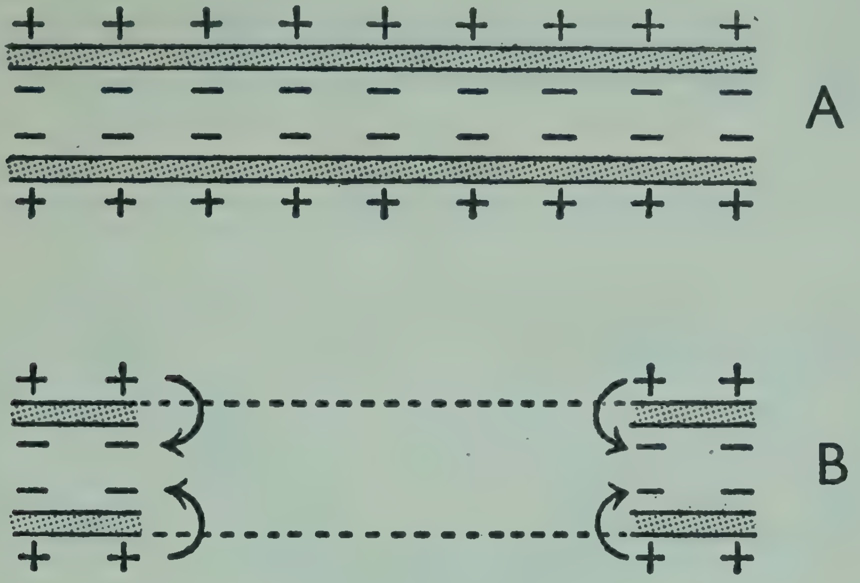


FIG. 41, 2

Older view of polarization and depolarization of a nerve fibre. Ions are indicated by the + and − signs. The arrows indicate the direction of flow of the current which by convention is that of the positive ions.

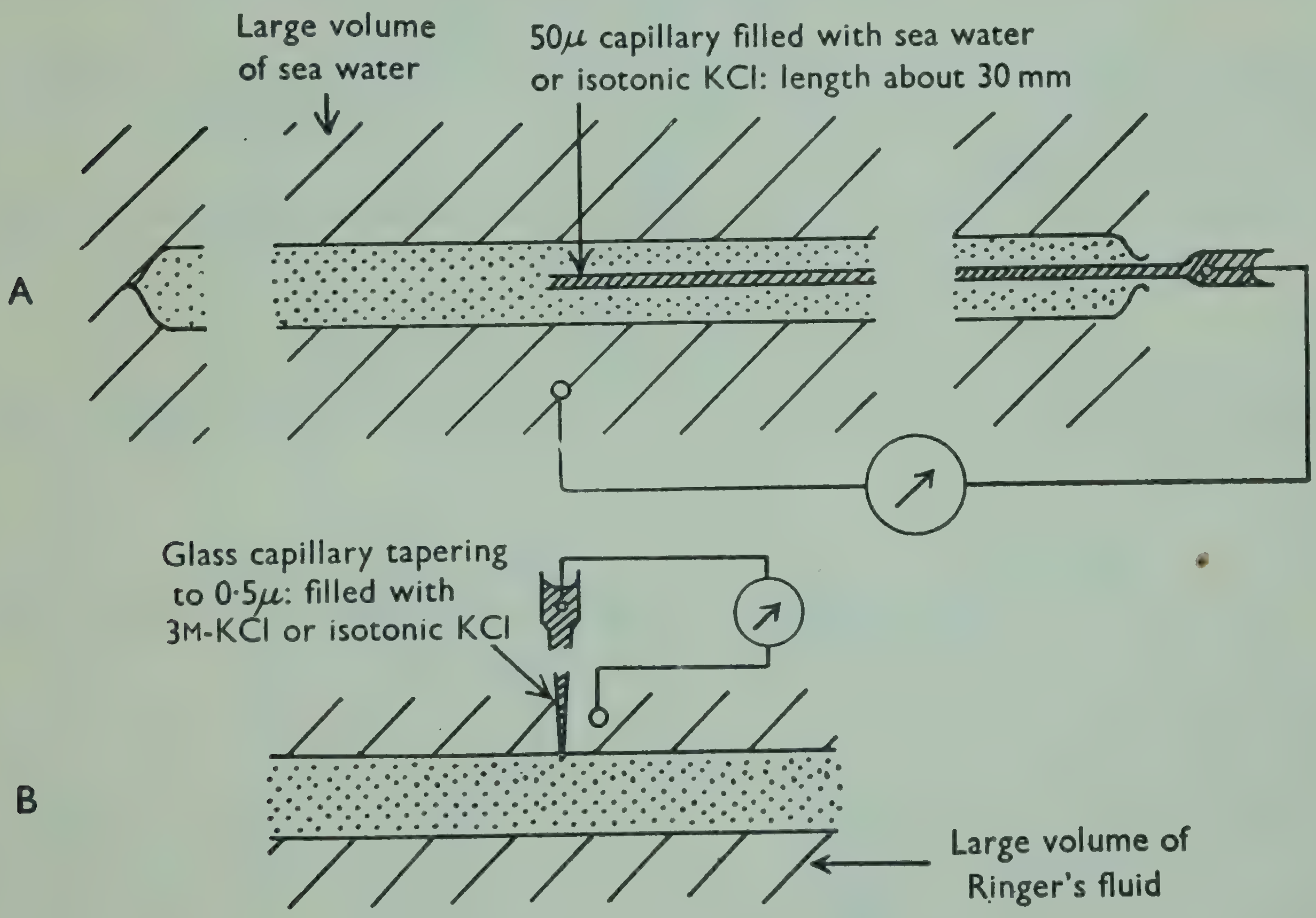

Large volume of sea water
50μ capillary filled with sea water or isotonic KCl: length about 30 mm
A
Glass capillary tapering to 0·5μ: filled with 3M-KCl or isotonic KCl
B
Large volume of Ringer's fluid

FIG. 41, 3

Methods for obtaining absolute values of resting potential and action potential. A, longitudinal insertion of internal electrode used for giant axon of the Squid, *Loligo*. B, transverse insertion of internal electrode used for muscle fibres and other cells (A. L. Hodgkin (1951), *Biol. Rev.* **26**, 342).

fast-conducting giant non-myelinated axons up to 1 mm in diameter which have proved particularly valuable in research into the mechanism of nervous conduction.

The nerve fibres within the brain and spinal cord have no neurilemmal sheath and this sheath may also be absent at the terminations of nerve fibres. The myelin sheath of fibres lying in the central nervous system is segmented but the nodes of Ranvier are not so easily demonstrated as in the peripheral nerves. On central fibres, 3 to 15 μ in diameter, the nodes are 200 to 2000 μ apart.

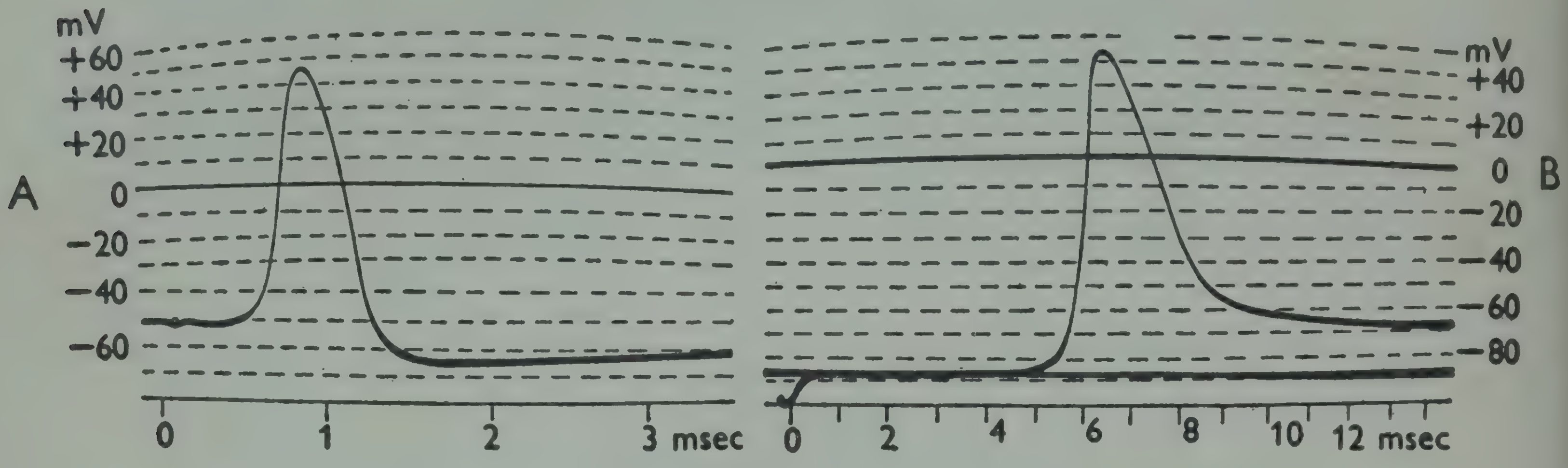


Fig. 41, 4

Resting potentials and action potentials recorded with internal electrodes. The ordinate gives the potential of the internal electrode relative to that of the external solution. The line through zero is the base line observed when the microelectrode was outside the fibre. The abscissa gives the time in msec. beginning at the moment of stimulation. A, *Loligo* axon 18·5° C obtained by the method shown in Fig. 41, 3, A. B, Frog muscle fibre from sartorius at 13° C obtained by the method of 41, 3, B (A. L. Hodgkin (1951), *Biol. Rev.* **26,** 343).

The axon of the typical motor cell passes without interruption to the muscle fibre, where it has a special termination (Fig. 42, 2). The axon may thus be anything up to several feet long, as in the case of the fibres going to the muscles of the hands and the feet. On the other hand many axons are quite short, for example those of the internuncial neurones (interneurones)

Table 41, 5

Concentrations of Ions in Frog Muscle and Ringer's Solution together with the Corresponding Membrane Potentials calculated from the Formula given on page 37.

Ion	Myoplasm m.equiv per l.	Ringer's Solution m.equiv per l.	Potential E mV
K^+	125	2·5	−98
Cl^-	1 to 5	117	−117 to −80
Na^+	15	117	+52
Anions	135		

After A. L. Hodgkin (1949), *Brit. med. Bull.* **6,** 322.

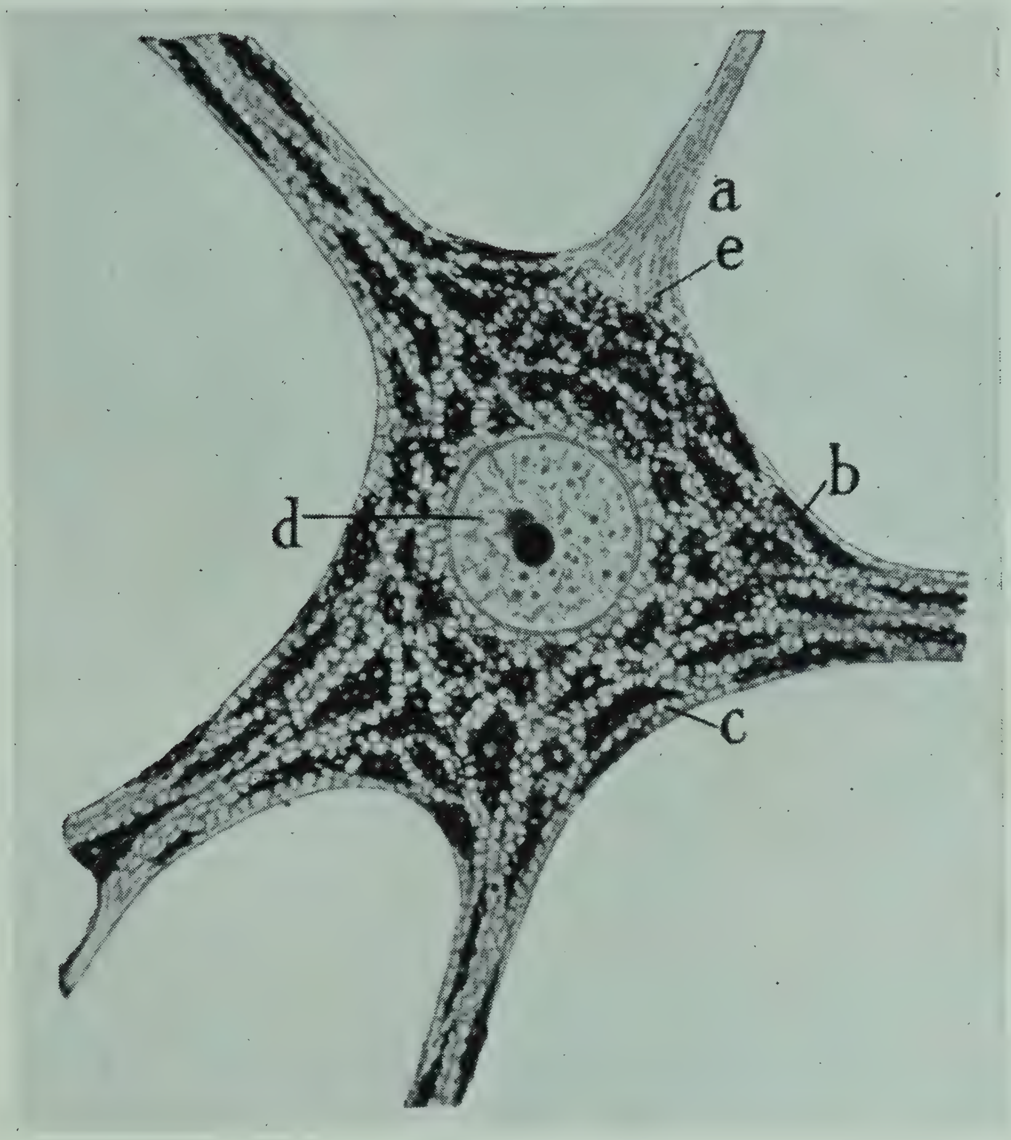


FIG. 41, 6

Motor cell from the spinal cord of the rabbit. (Thionin stain.) a, axis-cylinder; b, Nissl granules; c, cytoplasm; d, nucleus; e, axon hillock. (*After* Cajal (1933), *Histology*, p. 298. London: Baillière, Tindall & Cox.)

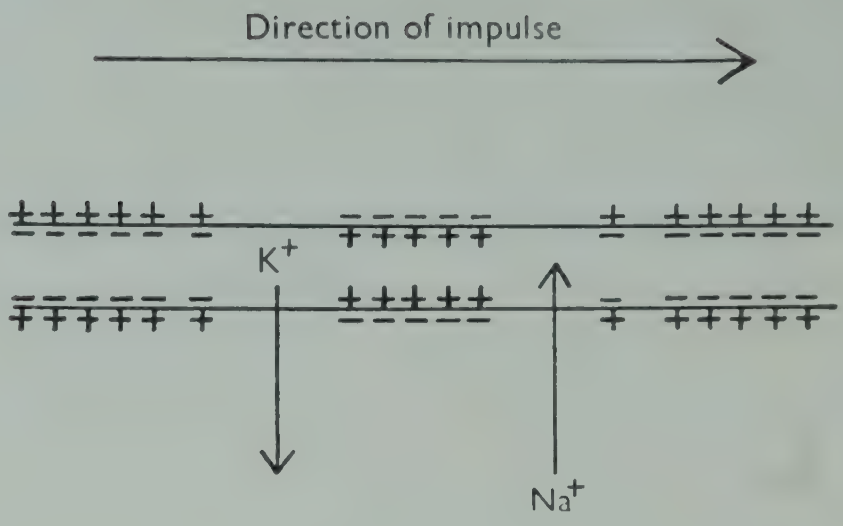

Direction of impulse
K+
Na+

FIG. 41, 7

Diagram to illustrate the ionic exchanges and the reversal of polarity of the membrane during a propagated action potential. There is first a transient increase in sodium permeability, then a longer increase in potassium permeability (A. L. Hodgkin (1951), *Biol. Rev.* **26**, 379).

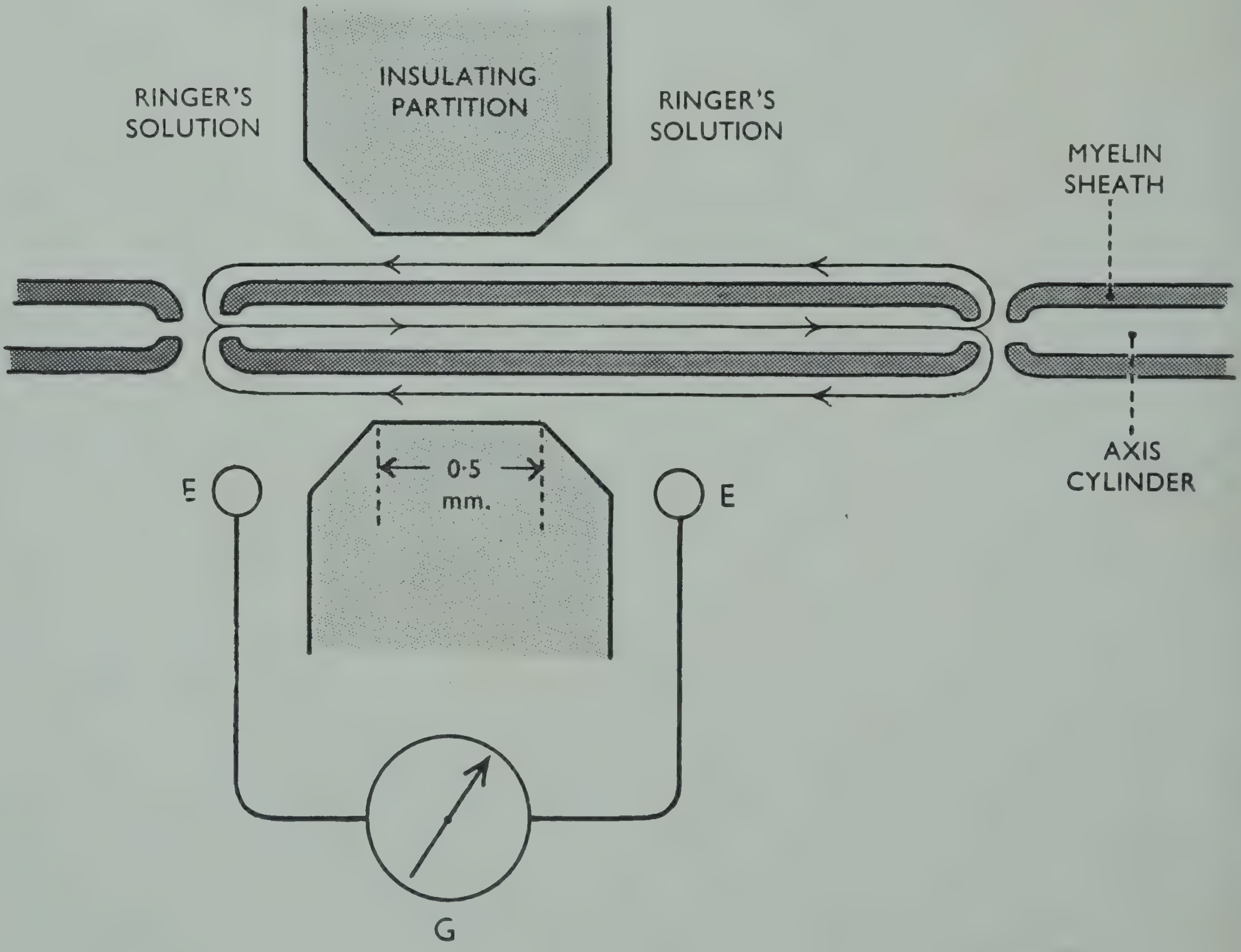

RINGER'S SOLUTION
INSULATING PARTITION
RINGER'S SOLUTION
MYELIN SHEATH
E
0·5 mm.
E
AXIS CYLINDER
G

FIG. 41, 8

Diagram of the method used by Huxley and Stämpfli to investigate saltatory conduction in nerve. The nerve fibre is drawn through a fine hole about 40μ in diameter in an insulator by a micromanipulator. The current flows along the axis cylinder out of one node and in at the other node as indicated by the arrows. The resistance of the gap between the insulators being about half a megohm, a potential difference set up between the two pools of Ringer's solution is measured by the galvanometer G connected to the two electrodes, one in each pool of Ringer's solution. The internodal distance in frog's myelinated fibres is 2 mm. (*After* A. F. Huxley & R. Stämpfli (1949), *J. Physiol.* **108**, 315.)

which make short connexions between nerve cells in the spinal cord. The peripheral nerves seen in an anatomical dissection are made up of hundreds of nerve fibres bound together into small bundles by fibrous tissue (perineurium); a second layer of fibrous tissue, the epineurium, binds the bundles together.

An analogy is often drawn between an axon and a telegraph wire. Both carry messages in their own special codes, but the analogy breaks down once that has been said because the physical properties of the two systems are so different. For example, the nerve impulse travels with an easily measured velocity and the electrical disturbance is confined at any one instant to a short length of the nerve; on the other hand an electric current in a wire can be detected in any part of the wire as soon as it starts to flow. The nerve fibre is often compared, with a nearer approach to accuracy, to a train of gunpowder: if this is set on fire at one end a flame passes along by igniting in succession the portions of the train immediately ahead. The nerve impulse is, like the flame, self-propagating.

The axon at rest is bounded by a polarized membrane as in Fig. 41, 2, A, and it has been supposed, until recently, that when a nerve is stimulated there is a movement of ions as in Fig. 41, 2, B, by which the potential across the membrane falls to zero. Recent investigations, mainly by Hodgkin at Cambridge, show that the original description must be elaborated somewhat.

The potential across the membrane of a living cell, whether nerve or muscle, can be measured by inserting into it a micro-electrode (Fig. 41, 3, A and B) which is connected to a valve amplifier and cathode-ray oscillograph. If the electrode is small enough it causes no apparent damage since the cell continues to function and its action potential is undiminished. The cathode-ray oscillograph tracings in Fig. 41, 4 show that the resting potential of squid nerve is of the order of -50 mV and of frog muscle about -90 mV, the interior of the cell being negative to the exterior. When the nerve is stimulated the membrane potential changes from -50 to $+40$ mV, a total change of 90 mV, the internal electrode being actually reversed in polarity for a short period. In the case of frog muscle the resting potential is on the average 88 mV, the mean value of the spike is 119 mV. In every excitable tissue that has been investigated with micro-electrodes, and the list now includes mammalian muscle and nerve, invertebrate nerve cells and vertebrate nerve cells in spinal cord and central nervous system and electroplates of *Electrophorus*, the sizes of the resting and action potentials are of the same order and reversal of potential at the peak of the spike is found although the forms and durations of the spike may be rather different. For example the duration might be 0·5 msec in a mammalian nerve fibre and 1000 msec in a frog cardiac muscle cell.

Analysis of nerve axoplasm and muscle cells shows that the concentration of potassium within the cell is much higher than that in the external fluid, as shown in Table 41, 5. On the other hand there is relatively little sodium or chloride within the cell. Since both nerve fibres and muscle cells of amphibia lose potassium and gain chloride when placed in Ringer's solution the resting membrane must be permeable to potassium and chloride; both the potassium and the chloride ions are moving from a zone of high concentration to one of low concentration. It is found that the ionic equilibrium of resting muscle and nerve can be expressed by a Donnan equation on the assumption that the principal ions concerned are sodium, potassium and chloride and that the membrane is impermeable to the sodium ions but permeable to the others. This leads to the relation

$$\frac{[\mathrm{K}]_i}{[\mathrm{K}]_o}=\frac{[\mathrm{Cl}]_o}{[\mathrm{Cl}]_i}=50$$

where $[\mathrm{K}]_i$ and $[\mathrm{Cl}]_i$ are the concentrations of the potassium and the chloride ions in the water in the cell and $[\mathrm{K}]_o$ and $[\mathrm{Cl}]_o$ are the corresponding activities in the external solution. The potential E_k across such a resting membrane is given by the formula on p. 37 which simplifies to

$$\mathrm{E}_k=58\log_{10}\frac{[\mathrm{K}]_i}{[\mathrm{K}]_o}=58\log_{10}50=98\ \mathrm{mV}.$$

This calculated value is not far from the resting value for muscle obtained experimentally. As predicted by this theory, increase of the potassium concentration in the external fluid reduces the membrane potential. The potassium and chloride system can be thought of as a battery which provides fairly large ionic currents during electrical activity. Experiments with radioactive ions show that the resting membrane is not, in fact, totally impermeable to sodium, and the inward and outward fluxes of sodium ions *at rest* are about the same size as the potassium ion fluxes *at rest*. However the net inward flux of sodium is about zero, since an active outward transport of sodium balances the inward leakage, so that the Donnan equilibrium relation still applies.

The records in Fig. 41, 4 show clearly that the membrane potential is reversed when the cell is active. If the membrane during activity became highly and selectively permeable to sodium these ions could then enter the fibre and reverse the potential. The calculated value (+52 mV in Table 41, 5) agrees reasonably well with the value +45 mV at the peak of the spike in Fig. 41, 4. If the depolarization of a nerve or muscle during activity is due to entry of sodium ions then cutting off the supply of sodium by immersing the cell in a sodium free medium should abolish excitability and prevent the conduction of an impulse. In fact, if the ions have free access to the surface of the cells conduction fails in a few seconds in a sodium free solution, the isotonicity of which is maintained by sucrose or glucose; conduction is restored by returning to Ringer's solution with normal sodium content. Reduction of sodium in the external fluid affects the resting membrane potential very little since the resting membrane is relatively impermeable to sodium. However, reduction of external sodium reduces the reversed potential as would be expected on the theory that there is a sudden appearance of sodium permeability when the cell is active and, further, the reversed potential is zero when the sodium concentrations inside and outside the cell are equal.

The minimum quantity of sodium, or rather the charge, which must enter the cell to produce the observed change of potential can be calculated from the known value of the capacity of the membrane. Since the amount which actually enters as judged by measurements using radioactive sodium is several times the calculated value, the sodium entry theory is upheld.

By the use of radioactive tracers the amount of potassium lost by nerves has been measured and in general the quantity of sodium which enters the nerve during one impulse is balanced by the amount of potassium vhich leaves it. If the sodium transfer and the potassium transfer ccurred simultaneously there would obviously be no net charge transferred cross the membrane and no action potential. The movements of these ons must, therefore, be separated in time. The action potential can be accounted for in both magnitude and sign if it is assumed that there is first a rapid transient increase in sodium permeability followed by a delayed but maintained increase in potassium permeability (Fig. 41, 7). The rapid entry of sodium ions in the rising phase of the action potential reverses the potential difference across the cell membrane; the local currents flowing into the active region of the nerve, by carrying a charge through the dielectric, depolarize adjacent regions of the membrane still in the resting state and make the impulse travel along the nerve. The increase in potassium permeability, over the resting permeability to that ion, becomes appreciable near the peak of the action potential and increases the rate at which these ions leave the fibre, so bringing the potential of the membrane back to its original resting value. The interchange of ions provides the energy for propagation of the nerve impulse and, during the period of recovery after the passage of the impulse, the interchange is reversed by a metabolic process. At the end of the action potential the permeability to potassium is increased for a little and there is a transient decrease in the permeability to sodium. These transient changes account for the refractory period (p. 745) since the first tends to keep the exterior of the membrane positive, as in the resting state, and the second prevents the entry of sodium when a stimulus is applied to the nerve.

These considerations suggest that when a nerve is stimulated electrically anywhere in its course one impulse should pass up the nerve and another down it from the point of electrical stimulation, and this actually does occur. An impulse which passes in the direction opposite to that normally taken is called an *antidromic impulse*.

In investigating the excitability and conducting properties of nerve it has been found most convenient to use electrical stimulation, because the duration and the intensity of the stimulus are easily controlled and repeated stimulation does not damage the nerve. When the stimulating current rises above a certain value a nerve impulse is initiated, but any increase in current beyond this value is usually without effect. There is thus "an all-or-nothing relation between the stimulus and the propagated disturbance" (Adrian) or, more briefly, the initiation of a nerve impulse is an all-or-nothing phenomenon. The value of stimulus which is just effective is known as the *threshold*. Sub-threshold stimuli cannot start off impulses but, for any particular nerve fibre, all stimuli greater than the threshold value initiate identical nerve impulses.

Hodgkin's theory of the activity in muscle and nerve provides an explanation for the existence of a sharp threshold. If a potential is applied across the membrane in the direction to depolarize it, there is an increase in the permeability to sodium. If the depolarization is small the current due to entry of sodium ions is smaller than the outward current carried by potassium and chloride ions; in these conditions the membrane repolarizes and no impulse is propagated. With slightly greater depolarization, however, a point is reached where the sodium entry becomes greater than the outward flow, so that the membrane potential continues to drop even when the stimulus ceases. This further increases the sodium permeability and hence, in a regenerative fashion, the rate of depolarization, so that there is a rapid inrush of sodium carrying the membrane potential towards its reversed value at the peak of the action potential and an impulse is propagated along the nerve.

Electrical stimulation of a nerve is more effective at the node of Ranvier than it is in the internodal region; also, blocking by polarization (see p. 754) and by narcotics is more effective at the nodes. These experiments suggest that excitation and other processes responsible for the conduction of the impulse occur mainly at the nodes and that the myelin is a shield or insulator which increases the conduction velocity by making the currents, produced during the activity of the nerve, act at a distance well ahead of the active region. In this process of saltatory conduction the nerve impulse is supposed to leap from node to node and not to progress in the continuous manner seen in non-myelinated fibres. To test this hypothesis Huxley and Stämpfli drew a nerve fibre through a glass capillary about 40 μ in diameter set in a partition between two troughs of Ringer's solution (Fig. 41, 8). The flow of current in the space around the nerve when it was active gave rise to a potential difference which was recorded by an amplifier and a cathode-ray oscillograph. If the current can enter or leave the axis cylinder only at the nodes of Ranvier, the current flowing along the axis

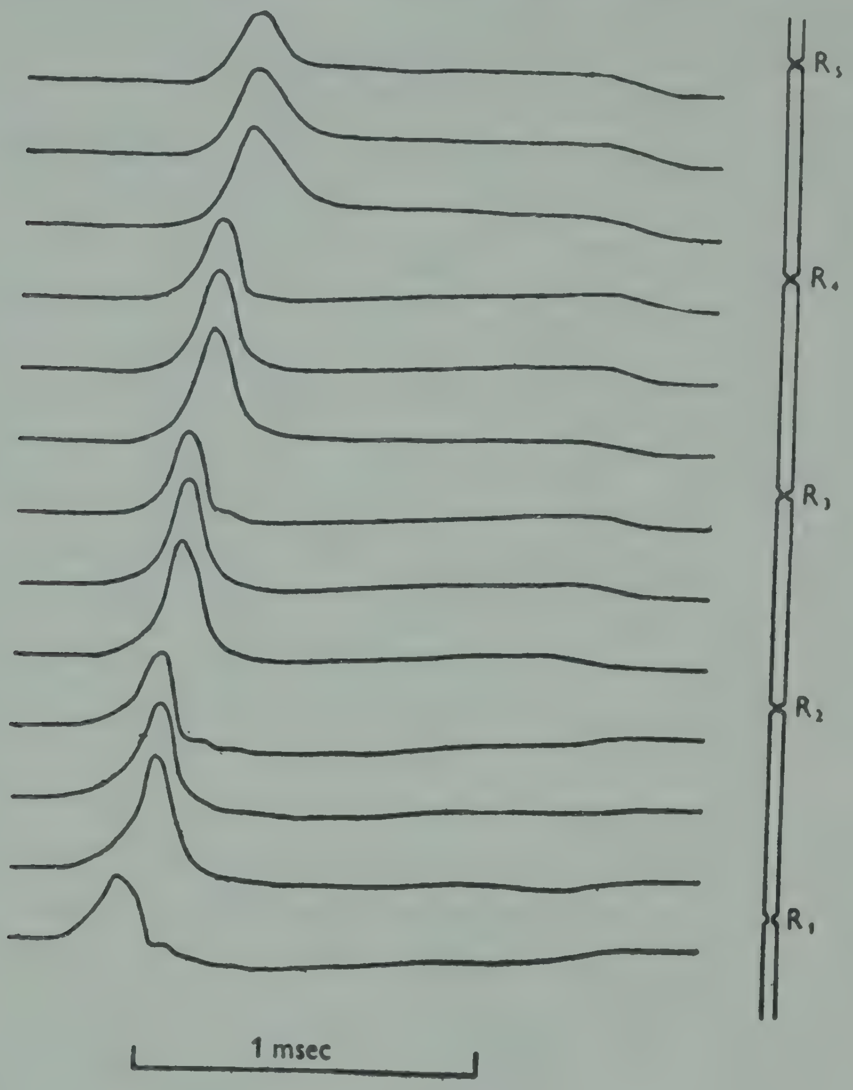


FIG. 41, 9

Records obtained by the method shown in Fig. 41, 8. Each cathode-ray oscillograph record is shown opposite the position on the fibre which the short glass capillary occupied when the record was obtained. All the records from any one internode are nearly the same and the peak of the disturbance occurs at the same time. R_1, R_2 to R_5 are notes of Ranvier (*After* A. F. Huxley & R. Stämpfli (1949), *J. Physiol.* **108,** 321.)

cylinder in the centre of the myelin sheath must be the same at all points n any one internode at any one instant and the internal current must be the same as the external current. The records show (Fig. 41, 9) that this s very nearly true, the discrepancy being accounted for by leakage of current through the myelin. The records also show that in any one nternode the peak of the electrical disturbance is reached at the same nstant; the peaks for any one internode are displaced stepwise in time from the peaks in the next internode. As the conduction time, that is the time from the application of the stimulus to the action potential, increases discontinuously, conduction in myelinated nerves can be described as saltatory.

That transmission of the nerve impulse depends on currents flowing outside the myelin sheath is shown by paring down the nerve of a nerve-muscle preparation till only one nerve fibre is left (Fig. 41, 10). Stimulation at the point P produced a visible contraction of a motor unit of the muscle M. The preparation was then laid on a pool of Ringer's solution on two insulated microscope slides A and B so that there was an air gap of 1 mm or so separating the slides with an internode, not a node of Ranvier, in the gap. Stimulation at A produced a muscle twitch if, and only if, there was a wet thread T making an electrical connexion between the two pools of Ringer's solution. In other words conduction is blocked if, by evaporation of water from the nerve fibre in the air gap, the external resistance around the nerve is above a certain value. The action potential at each node excites the next node by current flowing forwards in the axis cylinder and back in the fluid outside the myelin sheath as in Fig. 41, 8.

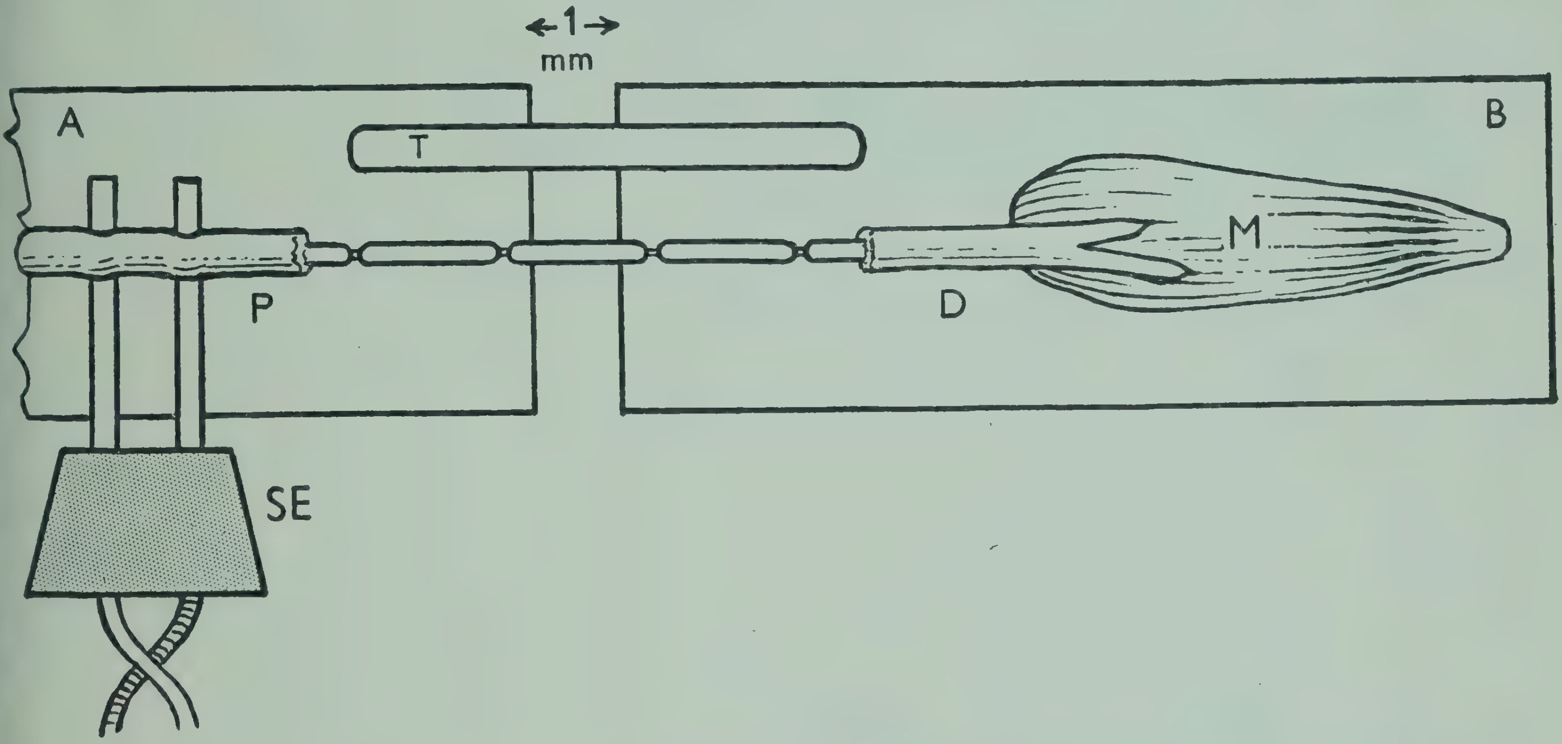


Fig. 41, 10

Diagram to illustrate the experiment described by A. F. Huxley & R. Stämpfli (1949), *J. Physiol.* **108,** 315, to show the importance for conduction of the nerve impulse of currents flowing outside the myelin sheath. A and B, microscope slides; P and D, proximal and distal portions of the nerve to the muscle M; T, wet thread; SE, stimulating electrodes.

Before the invention of microelectrodes all investigations of nervous activity had to be made by electrodes applied externally. This method is still widely used because in many experiments the information which is desired can be obtained by this relatively simple technique. We go on, therefore, to describe the records obtained by the use of external electrodes applied to nerve.

If electrodes applied to a nerve are connected to a galvanometer, as in Fig. 41, 11, A, no current flows when the nerve is not conducting impulses since all parts of the exterior of the nerve are at the same potential. When a brief electric current (or 'shock') of sufficient strength is applied to the nerve from the stimulating electrodes SE, a nerve impulse passes along the nerve, reaching successively the positions 1, 2, 3, 4, 5, at each of which depolarization occurs in turn. When the impulse reaches position 2, the electrode at 2, since the membrane is depolarized, is virtually in connexion with the negative interior of the fibre while the electrode at 4 is still in contact with the positively charged exterior of the polarized membrane; thus a current flows for a short time in the galvanometer from right to left. As the impulse passes along to position 4 these conditions are reversed, so that the galvanometer deflexions are recorded as a diphasic curve (Fig. 41, 11, B). Since the impulse travels rapidly along the fibre these deflexions are of short duration.

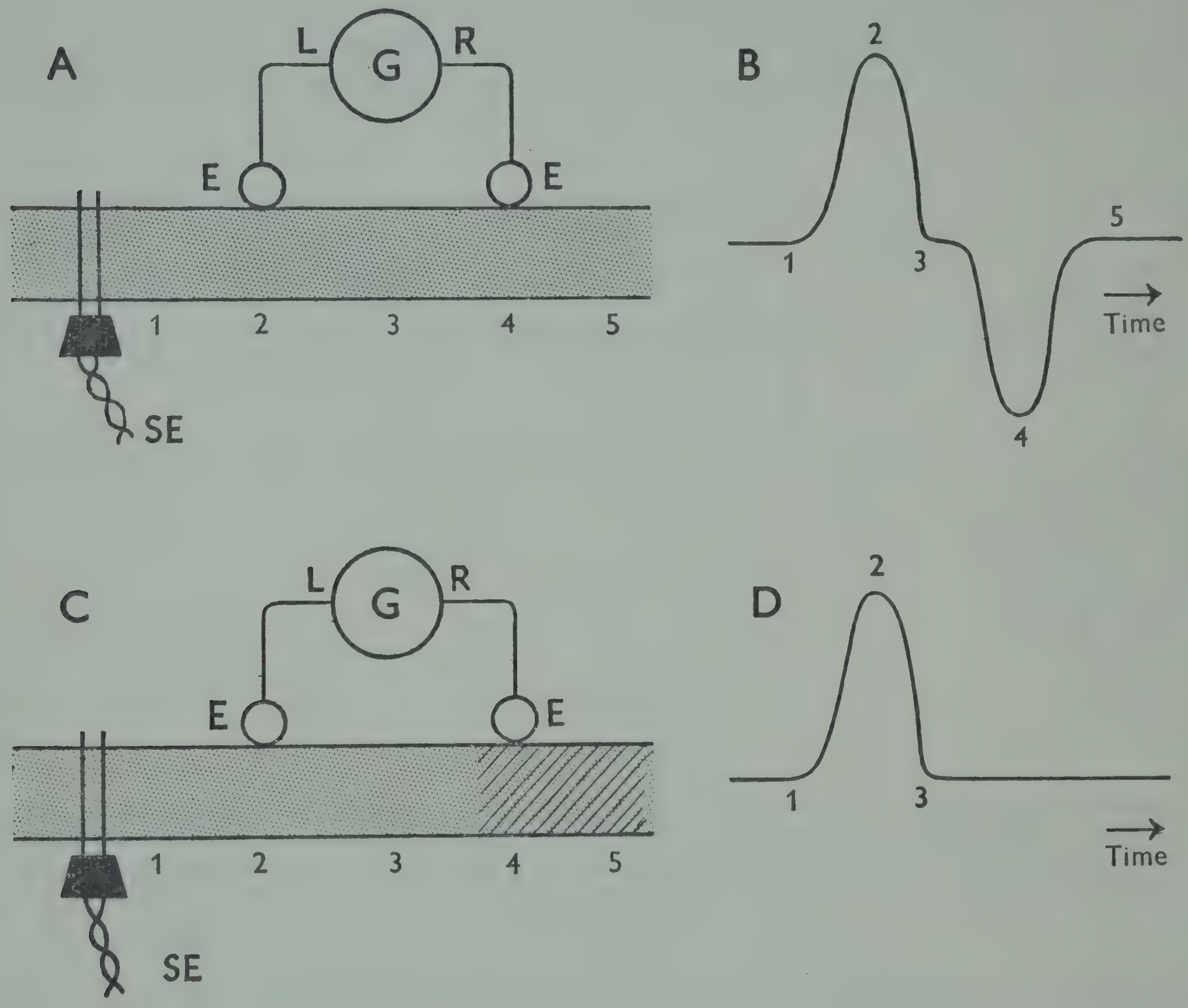


FIG. 41, 11

Diagrams illustrating the electrical changes accompanying the passage of a nerve impulse as revealed by externally applied electrodes. G, galvanometer; L and R, left and right; SE, stimulating electrodes; EE, electrodes resting on the nerve. The shaded area in C indicates that the nerve at that region has been killed. An upward deflexion in B and D, which are records of the galvanometer deflexions against time, indicates relative negativity of the electrode connected to L.

If regions 4 and 5 of the nerve are injured by applying a hot wire, or by dipping them in hot water (shaded area of Fig. 41, 11, C), the polarized membrane is destroyed. The electrode at 4 is now virtually in contact with the interior of the fibre while the electrode at 2 remains connected with the undamaged exterior of the fibre (Fig. 41, 11, C). Thus, even when the fibre is not conducting impulses, a steady current flows through the galvanometer from left to right. This *current of injury* produces a potential (*demarcation potential*) of as much as 30 millivolts in a freshly injured frog sciatic nerve; it gradually declines and disappears in a few hours. The potential difference is produced, as already explained, by the different concentration of electrolytes on the two sides of the nerve membrane.

When the nerve is stimulated at SE (Fig. 41, 11, C) the record of the current in the galvanometer due to the passage of the nerve impulse is monophasic (as in Fig. 41, 11, D) since no variation of potential can occur in the dead tissue under the electrode at 4. These monophasic records being simpler are more easily interpreted than diphasic records (Fig. 41, 11, B) and it is quite usual practice to record potentials between a normal and an injured point on a nerve.

The minute and rapid changes in potential which occur in nerve fibres during the passage of an impulse are known as *action potentials* or *spike potentials*. These provide only a few millivolts across externally applied electrodes and often the shunting effect of surrounding tissue reduces the spike to a few microvolts. The action potentials must be amplified by thermionic valves to produce a deflexion of the beam of a cathode-ray oscillograph which can be photographed. Such an oscillograph has the advantage of reproducing faithfully alterations in potential which occur too fast to be followed by a mechanical device

When a frog's sciatic nerve is very strongly stimulated all the nerve fibres in it are excited. Electrodes placed near the point of stimulation show the action potential as a simple wave but if the monophasic action potential is recorded from electrodes placed at a considerable distance from the stimulating electrodes a complicated wave form (Fig. 41, 12a) is obtained in which three groups of waves (spikes) labelled A, B and C can be recognized. A is further subdivided into α, β and γ. Each part of the complex wave of Fig. 41, 12, is contributed by a different group of nerve fibres; the nerve impulses all start from the point of stimulation at the same time but travel at different rates. Such a set of impulses initiated at the same instant is called a *volley*. In this particular experiment the distance from the stimulating electrodes to the recording electrode was 13·1 cm. If the time for the foot of the wave to reach the recording electrode is read off the logarithmic abscissa of Fig. 41, 12b, it is easy to calculate that the rate of conduction (in frog nerve at room temperature) was for α 41 metres per sec, for β 22 metres per sec, for γ 14, for B 4 and for C 0·7. The conduction velocities in mammals are somewhat faster, α 100, β 60, γ 40, B 10 and C 2 metres per second, but this difference can be adequately accounted for by the higher body temperature.

Many characteristics of a nerve fibre are related to its diameter. It has been found that the conduction rate in a large fibre is greater than in a small one. Since the fibres in a frog's sciatic nerve measure from 1 to 20 μ in diameter it is not difficult to understand how the conduction rates vary through the range 0·7 to 41 metres per sec. Furthermore the threshold voltage required to stimulate a large fibre is less than that required for a smaller fibre. In consequence of this the complexity of the electrical record of a volley depends on the strength of the stimulus. If the stimulus is weak only the α wave is seen; if it is stronger then both α and β waves are seen, and so on.

There is a continuous range of fibre sizes in any nerve, but it has not been possible to assign with any confidence particular functions to particular sizes of fibres. The sensory root of the spinal cord contains fibres giving A (that is α, β and γ waves), B and C waves; the motor root yields α and C waves, the latter going into the white ramus. The grey ramus, containing fibres passing from the sympathetic system, shows B and C waves. The fastest fibres (α) are motor fibres or afferent fibres conveying impulses from the muscles and skin. At least some of the fibres of the B and C groups convey pain impulses. The C fibres are probably nonmyelinated; they are found especially in autonomic postganglionic trunks.

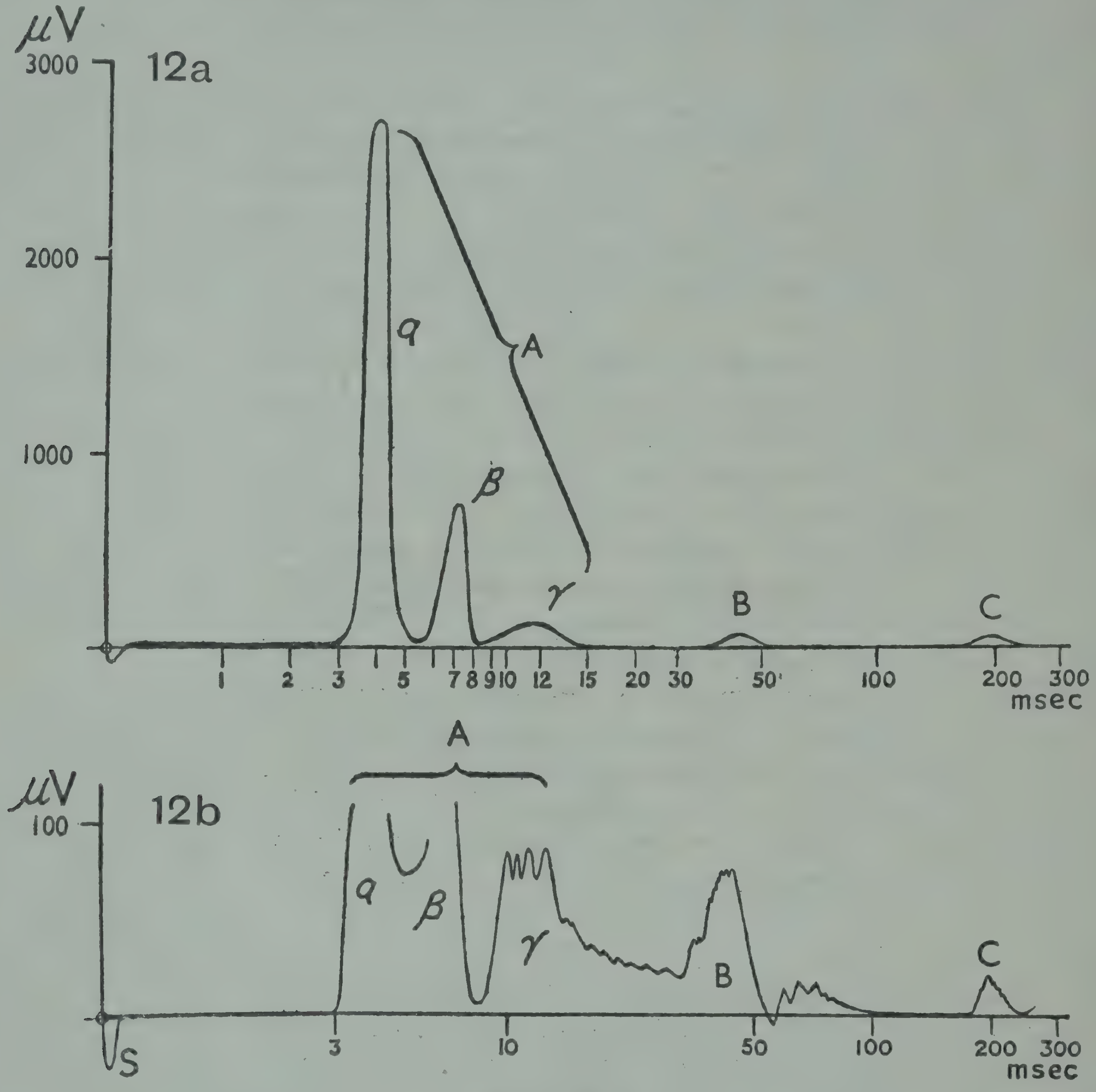


FIG. 41, 12

Complex action potential of frog's sciatic nerve led off at 13·1 cm from point of stimulation. The time is in milliseconds (logarithmic scale). Fig. 12b is at ten times the amplification of Fig. 12a. S = stimulus artifact, marking the zero point. (Drawn from diagrams given by J. Erlanger & H. S. Gasser (1937), *Electrical Signs of Nervous Activity*. Philadelphia: University of Pennsylvania Press.)

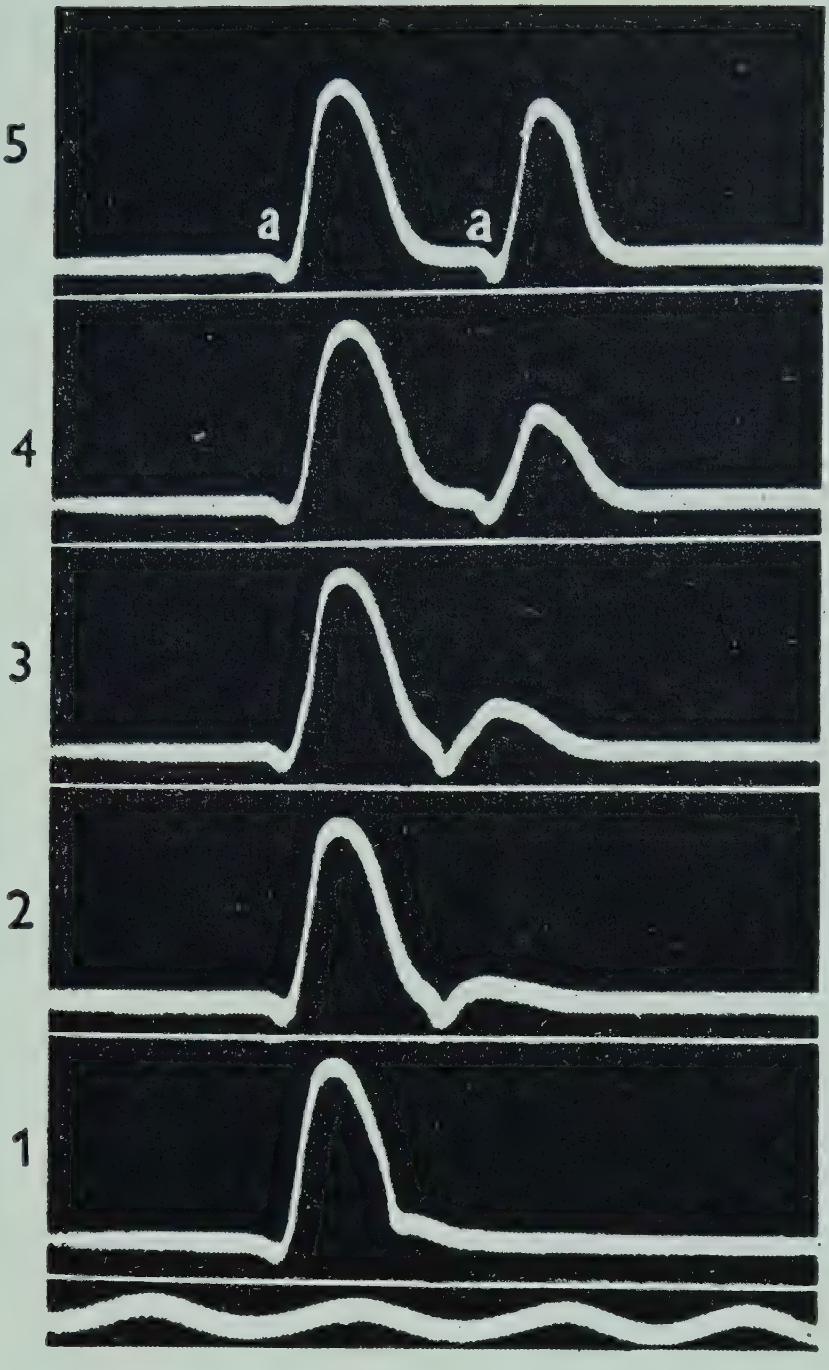


FIG. 41, 13

Relatively refractory period of A fibres of the phrenic nerve mapped in terms of the height of the response evoked by a supramaximal testing shock. The separation of the conditioning and testing shocks can be determined from the shock artifacts, a, a. At the interval of the lowermost record (about 0·5 msec) the nerve is completely refractory, but a small response is present in the next record. Conduction distance, 8 mm. Temperature, 38° C. The sine wave at the bottom of the figure gives time in msec (J. Erlanger & H. S. Gasser (1932), *Electrical Signs of Nervous Activity*, p. 172. Philadelphia: University of Pennsylvania Press).

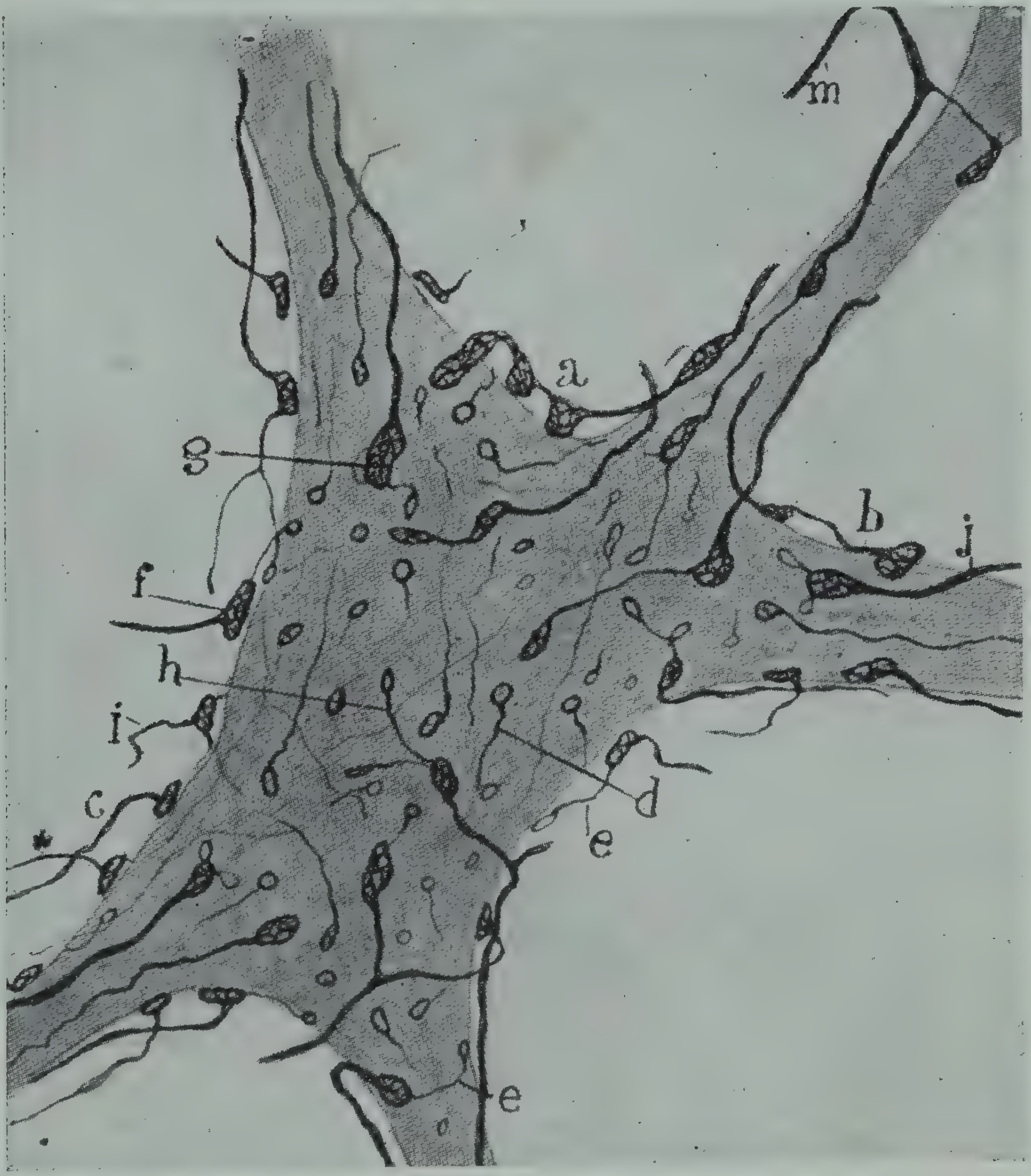


FIG. 41, 14

This drawing of a motor cell of the spinal cord shows that it is invested by a large number of synaptic knobs from many neurones. a, b, strong bulbs; c, i, terminal buttons; d, e, fine bulbs; m, terminal nerve fibre (Cajal (1933), *Histology*, p. 290. London: Baillière, Tindall & Cox).

The behaviour of single axons can be studied by cutting partly through a nerve until only one axon is left intact. The action potential, or spike, in a single axon is found to be an all-or-nothing phenomenon, its size depending on the prevailing conditions, for example temperature, and not on the intensity or manner of stimulation. The spike is immediately followed by a negative after-potential lasting about 10 milliseconds. This is succeeded by a positive after-potential lasting about 50 msec. The after-potentials are only a small fraction of the spike potential (Fig. 41, 15) and are not seen unless the valve amplifier has considerable magnification and can amplify slow changes of potential. The after-potentials are the electrical signs of recovery processes, being increased when the metabolic rate is increased and diminished when the oxidative processes are diminished.

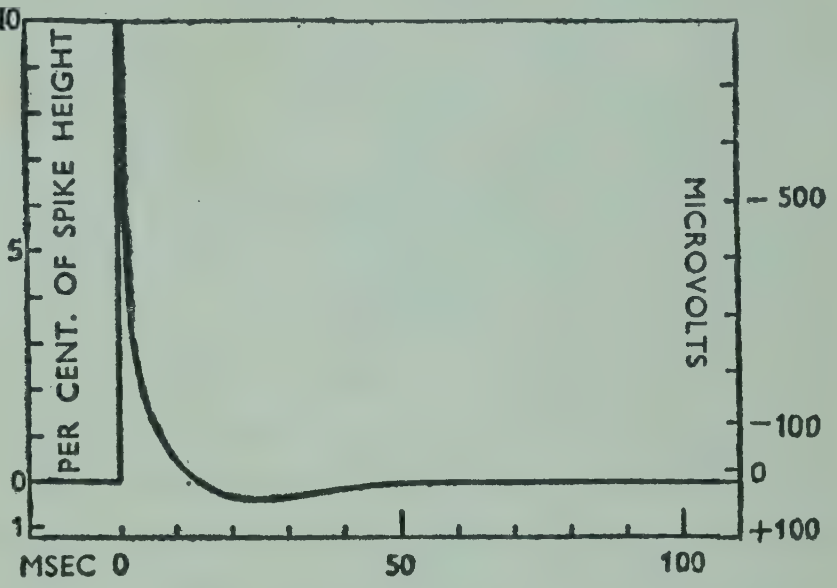


Fig. 41, 15

Phrenic nerve at 37° C. The spike, lasting only 0·5 msec, leaves the base-line and goes above the figure; it then returns to the figure and is continued for about 10 msec as the negative after-potential, which is succeeded by a positive phase, the dip below the base line, lasting about 50 msec. (J. Erlanger & H. S. Gasser (1937), *Electrical Signs of Nervous Activity*, p. 148. Philadelphia: University of Pennsylvania Press).

It has been known for a very long time that if a nerve is stimulated twice in quick succession it may not respond to the second stimulus. From oscillograph records it is found that the *absolutely refractory period* of mammalian A fibres is about 0·4 msec, which is also the approximate duration of the spike (Fig. 41, 13). During the absolutely refractory period no second shock, *however large*, can elicit a response. When the interval between two stimuli is increased (Fig. 41, 13, 2 to 5) the spike produced by the second shock gradually increases till at 1 msec (Fig. 41, 13, 5) nearly 90 per cent. of recovery has taken place; the second spike reaches the same height as the first when the two shocks are about 3 msec apart. Further, the threshold strength of shock required to elicit a second spike is raised from 0·4 msec onwards but falls to normal again in about 3 msec. Since excitability is usually expressed as the reciprocal of the threshold stimulus it can be said that the excitability rises to normal in about 3 msec after an effective stimulus. The time from 0·4 to 3 msec is called the *relatively refractory period.*

Since recovery of an A fibre is complete in 3 msec a single axon can carry over 300 full-sized spikes per second. In normal circumstances, however, it is seldom that the frequency of the impulses is over 100 per second (see p. 767). A nerve fibre can respond to artificial stimulation at the rate of 1000 per second but, since recovery is by no means complete after 1 msec, the spikes after the first are small and tend to decline during the course of the stimulation. The absolutely refractory period thus sets an upper limit to the number of impulses which a nerve can transmit in a given time.

During the relatively refractory period the spike size, the excitability and the conduction rate are all subnormal. A supernormal period follows and coincides in time-course with the negative after-potential; during this period the excitability, conduction rate and spike size are supernormal. During the positive after-potential the excitability and the conduction rate are subnormal but the spike is normal in height. The negative after-potential can be regarded as a residual depolarization and the positive after-potential as a hyperpolarization or anelectrotonus (see p. 755).

Excitable tissues such as nerve can be stimulated only when the applied voltage is changing and a constant current applied to a nerve is normally effective only at the opening and closing of the circuit. If the excitability is raised, however, a constant current can produce a series of spikes. When a nerve is injured by cutting, a persistent area of depolarization is created and this often sets up a so-called injury discharge, usually of fairly high frequency.

Long before oscillograph records of action potentials could be obtained the problem of measuring the speed of conduction of the nerve impulse was solved by Helmholtz (1852). His method is the basis of a well-known class experiment. The muscle of a frog's sciatic-gastrocnemius preparation is attached to a lever writing on a smoked drum. The nerve is first stimulated as far away from the muscle as possible; the latent period, that is the time from the moment of stimulation to the rise of the lever, is measured. This latent period includes the shock-response time of the nerve, the conduction time in the nerve, the delay at the myoneural junction, as well as mechanical delays in both muscle and recording lever. The latent period when the nerve is stimulated as near as possible to the muscle also includes all these delay times, but it is slightly shorter by the time taken for the impulse to pass between the two points of stimulation. The distance on the nerve between the points stimulated divided by the difference in the latent periods gives the speed of the nerve impulse. The speed found in this way is, of course, the average speed of the predominant size of nerve fibre.

THE SYNAPSE

The meeting-place of two neurones is called a synapse but there is no actual continuity between one neurone and the next. The histological picture of a synapse varies but usually an axon near its termination breaks up into numerous fine branches with little terminal swellings, *synaptic knobs*, which lie close to the body or dendrites of the next neurone. A single anterior horn cell may be invested by as many as 1300 synaptic knobs (Fig. 41, 14) derived from a large number of axons.

Many of the investigations on the synapse have been carried out on the superior cervical ganglion because records of potential changes arising in it can be made without disturbing its blood supply. When the electrical connexions are arranged as in Fig. 41, 16, B, the spike potentials S_1 and S_2 are seen after a single shock sent in at SE. When a lead is taken from the ganglion as in Fig. 41, 16, A, then the spike potentials are followed by a slow negative wave N and a slow positive wave P; both of these waves must therefore arise in the ganglion cells. If the postganglionic fibres are stimulated through electrodes placed on them, impulses are propagated away from the electrodes in both directions. The antidromic impulses arrive at the ganglion and potentials can be detected there, but impulses cannot be set up in the preganglionic fibres. In other words, impulses cannot be transmitted backwards across the synapse.

When the preganglionic trunk is stimulated impulses are conducted along it to the ganglion at rates which can be measured by the method given on p. 743. The postganglionic spikes occur somewhat later in the records than would be expected by calculation from these velocities and the distances to be covered; the delay presumably occurs in the synapse. This delay in the case of the sympathetic ganglion varies from 3 msec in the fastest fibres to about 30 msec in the slowest fibres.

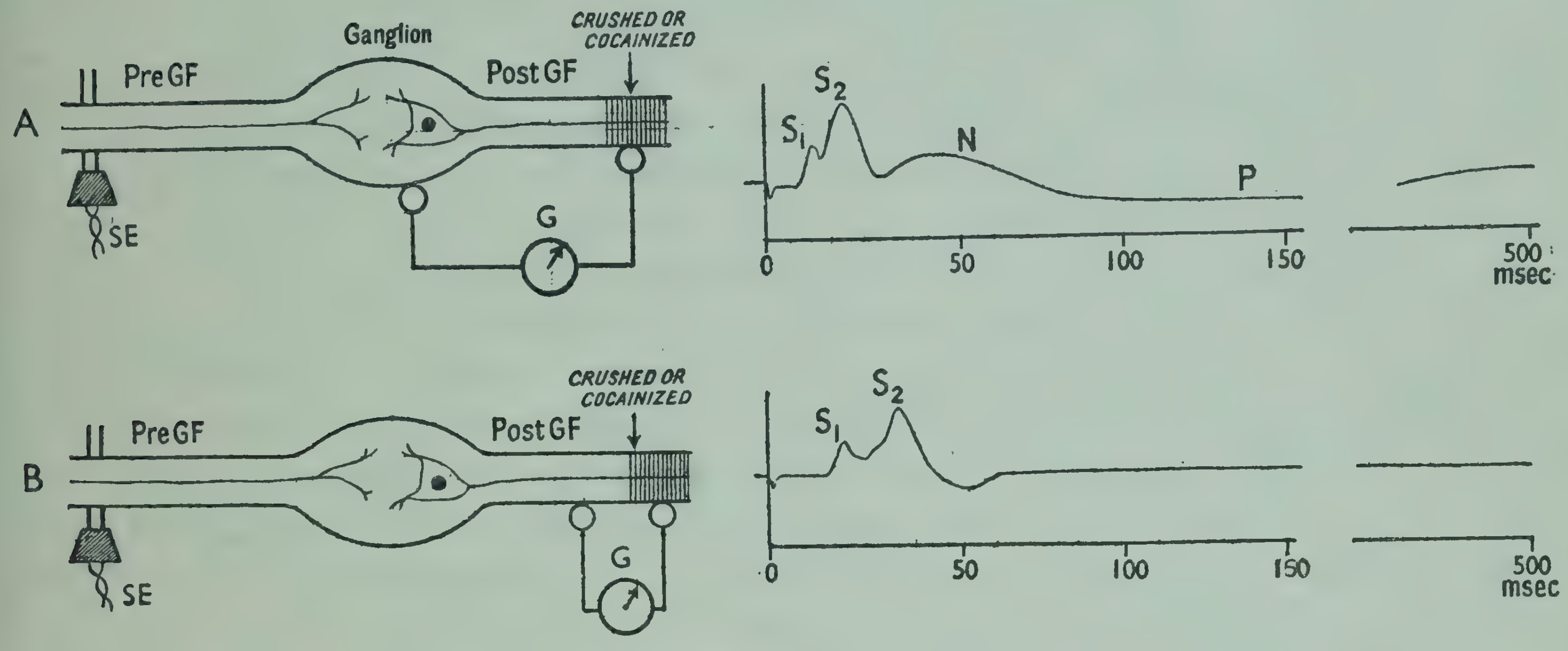


FIG. 41, 16

This diagram shows the method of studying the potentials produced in a sympathetic ganglion and in the postganglionic fibre (Post GF). SE, stimulating electrodes; PreGF, preganglionic fibre; G, galvanometer. With the galvanometer electrodes in position A the upper record on the right is obtained, when the arrangement is as in B, the lower record is made. S_1 and S_2 are spikes due to fast and slow conducting fibres respectively. N, slow negative ganglion potential; P, slow positive ganglion potential (J. C. Eccles (1935), *J. Physiol.* **85,** 500).

The refractory period of the synapse can be measured by sending in two stimuli in quick succession to the preganglionic fibre as in the case of the nerve fibre (Fig. 41, 13). In the synapses of the fastest fibres the absolutely refractory period is about 2·4 msec and the relatively refractory period is approximately 10 msec. If the interval between the stimuli is greater than this, say 10 to 50 msec, the response of the S_1 fibres to the second stimulus may be greatly increased or 'facilitated' (S_1 fibres are those which produce the spike marked S_1 in Fig. 41, 16, A and B). The first preganglionic volley makes a number of ganglion cells discharge and send impulses along the postganglionic fibres; at the same time the preganglionic volley raises the excitability of neighbouring ganglion cells without causing them to discharge. This area of increased excitability is called the *subliminal fringe* (see also p. 832). The second preganglionic volley makes a large number of these easily excitable cells discharge. The state of raised excitability corresponds in time-course to the negative wave of the ganglion potential (N, Fig. 41, 16, A). If the two preganglionic shocks are separated by 50 to 600 msec it may be found that the S_2 fibres (slow-conducting vasoconstrictor fibres in this case) show a smaller response to the second shock. This depression of excitability corresponds in time-course to the positive wave (P, Fig. 41, 16, A).

The preganglionic trunk of the superior cervical ganglion is formed by the junction of two trunks which arise from the lower sympathetic ganglia and pass as the ansa subclavia on either side of the subclavian artery. If one branch of the preganglionic trunk is stimulated a second stimulus applied very shortly after to the other branch may produce only a small action potential in the postganglionic trunk. This phenomenon is called *occlusion*, and is due to the fact that many of the cells in the ganglion receive synaptic knobs from the fibres of both branches of the preganglionic trunk; if these cells are fired by impulses arriving by one trunk, they are for a time refractory to impulses arriving along the other trunk. If a ganglion is painted with 0·2 per cent. nicotine solution no action potentials are obtained in the postganglionic fibres after a preganglionic volley. Since the preganglionic impulses are not affected the block must occur earlier than the site of production of the propagated action potentials.

It is possible to perfuse the blood vessels going to the cervical ganglion and to collect the perfusate. If the perfusion fluid contains eserine, acetylcholine can be detected in the perfusate when the preganglionic fibres are stimulated. If sufficient acetylcholine is added to the perfusing fluid, the ganglion cells can be made to discharge; weaker concentrations of acetylcholine increase the excitability of the cells to preganglionic impulses. When high concentrations of acetylcholine are perfused the ganglion cells are first stimulated and then paralysed because they are depolarized. These findings suggest that acetylcholine is the transmitter of the effects of the nerve impulses across the synapses in this ganglion.

Acetylcholine is produced at the preganglionic nerve-endings. After degeneration of the preganglionic fibres it cannot be detected and at the same time the enzymes necessary for the formation and destruction of acetylcholine, namely choline acetylase and cholinesterase disappear from the ganglion. The ganglion cells are therefore not concerned in the metabolism of acetylcholine and are merely effector cells. After preganglionic degeneration they remain sensitive to acetylcholine.

The importance of the acetylcholine-choline esterase system in nervous tissue is so great that it is appropriate at this stage to mention the mechanism for the formation of acetylcholine. Nervous tissue, as already mentioned, contains the enzyme system *choline acetylase* which catalyses the formation of acetylcholine. It requires the presence of potassium and magnesium ions and catalyses the transfer of the acetyl group from acetylcoenzyme A (p. 316) to choline.

$$\text{choline} + \text{acetyl CoA} \longrightarrow \text{acetylcholine} + \text{CoA}$$

When acetylcholine is broken down by choline esterase free choline and acetate are liberated. In the presence of ATP as a donor of energy the acetate reacts with coenzyme A to form acetylcoenzyme A.

$$\text{acetate} + \text{CoA} \xrightarrow{\text{ATP}} \text{acetyl CoA}$$

Lack of thiamine disturbs the properties of the nerve fibre. It is slowly used up by nerve tissue and has to be replenished but its biochemical function is unknown.

Drugs which block transmission in autonomic ganglia are much used in clinical medicine and the preceding description shows how such a block might be brought about. As already mentioned, excess acetylcholine, by causing a prolonged depolarization of the ganglion cell, prevents it from responding to further release of acetylcholine; nicotine also has this effect. Secondly local anæsthetic drugs such as procaine may cause a block by preventing the release of acetylcholine. Block can also be produced by substances such as curare and hexamethonium which compete with acetylcholine released from the nerve endings for the receptor sites on the ganglion cell.

PROPERTIES OF SYNAPSES IN THE SPINAL CORD

The synaptic connexions in the spinal cord are complicated and many alternative paths are available. The stretch receptors of a muscle send impulses in large diameter axons through the posterior roots into the spinal cord where branching of the axons takes place. Some of these branches end directly on motoneurones which, if they discharge, cause contraction in the muscle from which the afferent impulses arose; others end on the motoneurones of synergic muscles and some connexions reach the motoneurones of antagonistic muscles.

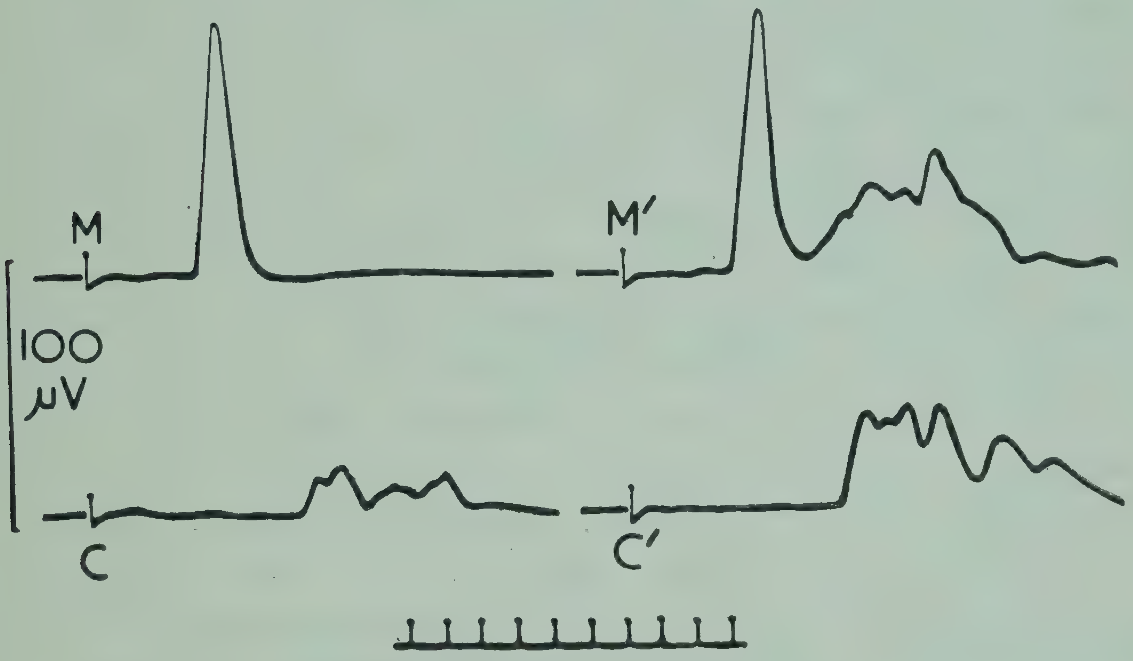


FIG. 41, 17

Records of potentials in the ventral roots of the spinal cord of a cat. Above after weak and strong, M and M′, shocks to a muscle nerve; below after weak and strong, C and C′, shocks to a cutaneous nerve. The anterior roots have been cut distal to the recording electrodes. Time in msec.

(*By courtesy of W. Schlapp.*)

A monosynaptic pathway is normally activated by stretching extensor or flexor muscles (indeed the myotatic reflex or tendon jerk in man and animals is brought about in this way); it can also be excited by electrical stimulation of the sensory nerve fibres of a muscle or by electrical stimulation of the dorsal root. It is possible to stimulate the large sensory fibres selectively by using a weak stimulus because they have the lowest threshold. The picture of the potentials in the ventral root (Fig. 41, 17 at M) shows first a stimulus artifact due to spread of stimulating current. This is followed after a latent period by a single spike potential. When the known times of conduction in the sensory and motor nerves are subtracted from the latent period, the time in the cord (central reflex time, p. 832) amounts to approximately 0·8 msec, but may vary slightly with the excitability of the motoneurones. It may be in part accounted for by reduced speed of conduction in the very fine presynaptic fibres ending in synaptic knobs but the greater part is due to synaptic delay.

Stronger stimulation of a posterior root or of a muscle afferent nerve may activate large and small fibres and the ventral root record (Fig. 41, 17 at M′) shows, in addition to the monosynaptic response, reflex discharges with a longer central reflex time, indicating the existence of several synapses in the pathway. Such a pathway is referred to as 'polysynaptic' and can also be activated by stimulating a cutaneous nerve, in which case no monosynaptic component appears in the response (Fig. 41, 17 at C and C′). As the efferent impulses which leave the spinal cord when a cutaneous nerve is stimulated can be traced only to the muscles of flexion, it can be said that the flexor reflex (see p. 830) is subserved by a polysynaptic pathway.

The monosynaptic pathway provides a simple means of investigating synaptic phenomena in the spinal cord. Thus facilitation may be demonstrated in the following way. If volleys of impulses coursing in the afferent nerves from two synergic muscles arrive at the motoneurones simultaneously, the ventral root response exceeds the sum of the responses to the individual volleys. This facilitation diminishes as the two volleys are separated in time up to about 15 msec, after which facilitation no longer takes place.

Inhibition may also be studied. When a monosynaptic response is elicited by stimulating the afferent nerve of a muscle, its magnitude can be reduced by previous stimulation of the afferent nerve of an antagonistic muscle and the motoneurones are said to be directly inhibited. This direct inhibition, which does not occur with simultaneous volleys, takes a small fraction of a millisecond to develop though, like facilitation, it lasts about 15 msec. One interpretation of this difference in the time course of direct inhibition and of facilitation supposes that the pathway for the former involves an interneurone.

Work by Eccles is beginning to lead to an understanding of these findings. He succeeded in inserting glass microelectrodes about 1 μ in diameter into the motoneurones of the cat's spinal cord so that the activity of a single motoneurone could be studied in virtual isolation from its neighbours. The resting potential (*cf.* p. 737) was about 70 mV; as the action potential was usually 90 to 100 mV there was a reversal of membrane potential of the order of 25 mV. Impulses arriving along monosynaptic pathways produced a brief partial depolarization (the synaptic potential) of the membrane of the motoneurone; when the depolarization, that is the reduction in membrane potential, was sufficiently large, about 10 mV, the neurone discharged and an impulse passed along the axon. The synaptic potential lasted several milliseconds and the synaptic potential produced by one volley could summate with that produced by a second to make the motoneurone discharge. Thus we have a satisfactory explanation for facilitation. Inhibitory impulses from fibres in antagonist muscles reaching a motor neurone produced a brief increase (about 2mV) in the membrane potential (hyperpolarization or anelectronus). This hyperpolarization was able to counteract the depolarization produced by excitatory impulses. Thus the same motoneurone responds to impulses arriving in one nerve fibre by depolarization and to impulses arriving in another fibre by hyperpolarization. Since it is not possible to explain these findings electrically, Eccles suggests that excitation is brought about by the release of a chemical substance which initiates a net inward flux of sodium ions whereas inhibition is due to the release of another chemical mediator which raises the permeability of the post-synaptic membrane to potassium and chloride ions in such a way as to increase the membrane potential slightly.

PROPERTIES OF THE SYNAPSES IN THE MID-BRAIN

When an afferent nerve is stimulated the resulting reflex muscular movement begins after a short latent period. This interval, which exceeds that calculated from the conduction velocities and the length of the nerve pathway in the reflex arc, represents the sum of the conduction time in the afferent nerve, the delay at the synapses in the cord, the conduction time in the outgoing nerve, the end-plate delay (that is the time elapsing between the arrival of a nerve impulse at the muscle and the setting up of the excitatory process in the muscle), and the conduction time in the muscle. All these, except the synaptic delay, can be measured directly. If the time of conduction in the afferent and efferent paths is subtracted from the observed latent period we have the *reduced latent period.* The length of this must obviously depend on the number of synapses within the central nervous system which are involved in the reflex.

Lorente de Nó has measured the synaptic delay in the mid-brain of the rabbit by the method indicated in Fig. 41, 18. Stimulating electrodes (SE) placed in the brain substance near the VIth nerve nucleus stimulated the VIth nerve directly and the IIIrd nerve indirectly by way of the fibres of the medial longitudinal bundle (MLB). The medial rectus (supplied by the IIIrd nerve) contracted later than the lateral rectus (supplied by the VIth nerve). Since the two pathways are of nearly identical length the time difference, 0·5 to 1·36 msec, must be due to synaptic delay. The delay varied in inverse relation to the strength of the stimulus.

By sending in two successive stimuli at SE it was found that the upper limit of the absolutely refractory period of the synapse was 0·6 msec, which is the same as the absolutely refractory period of the nerve fibre.

If the reticular substance in the pons and medulla which gives rise to the medial longitudinal bundle is destroyed then single shocks at SE (Fig. 41, 18) are ineffective, although two successive shocks may cause a muscle twitch. This is taken to indicate that the impulses from the reticular neurones, by raising the c.e.s., may have a facilitating effect on the neurones in the nucleus of the third nerve.

The period of facilitation after a shock can be investigated by noting the effect of a second shock sent in shortly after the first. Facilitation lasts from 1·5 to 80 msec according to the strength of the shock. This great variation in the period of facilitation can be explained on the basis

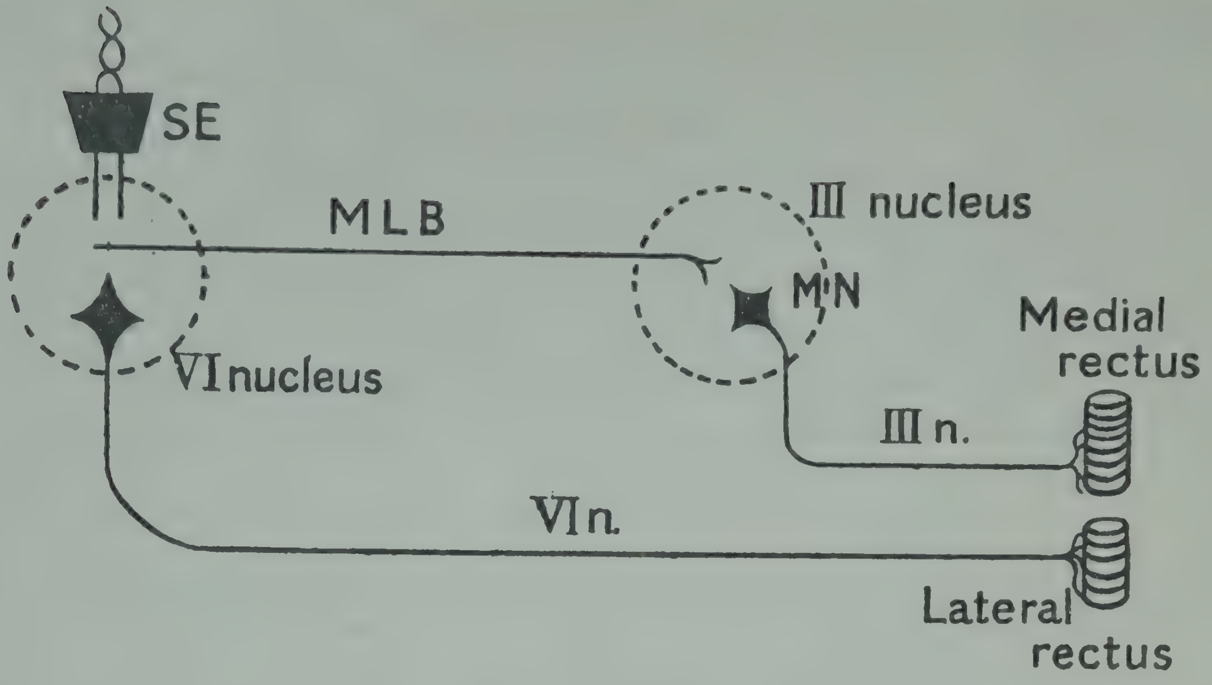

SE
MLB
III nucleus
MN
Medial rectus
VI nucleus
III n.
VI n.
Lateral rectus

FIG. 41, 18

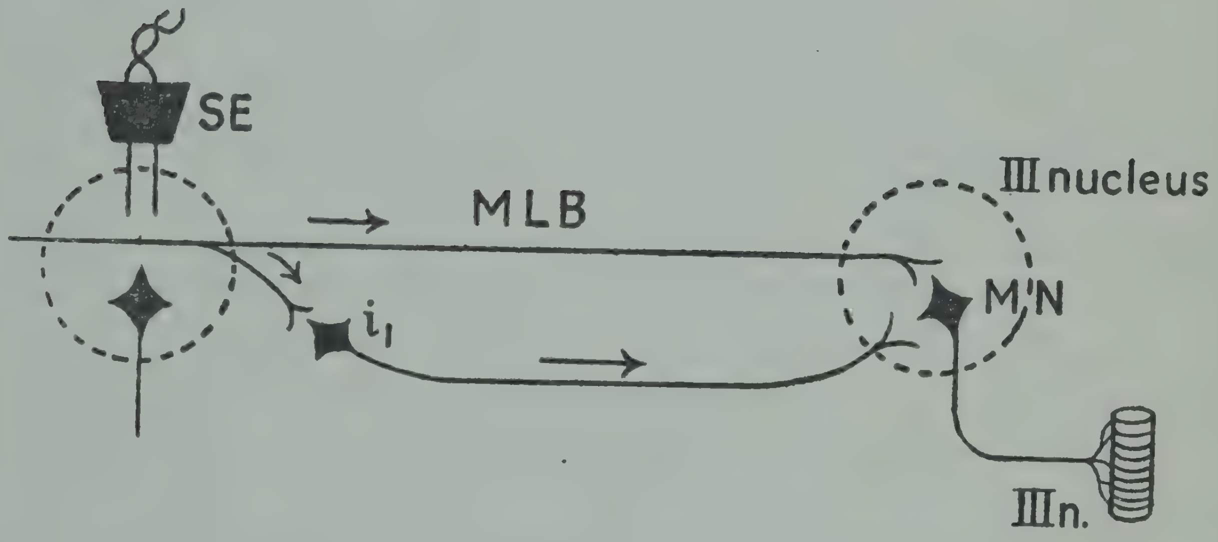

SE
MLB
III nucleus
i₁
MN
III n.

FIG. 41, 19

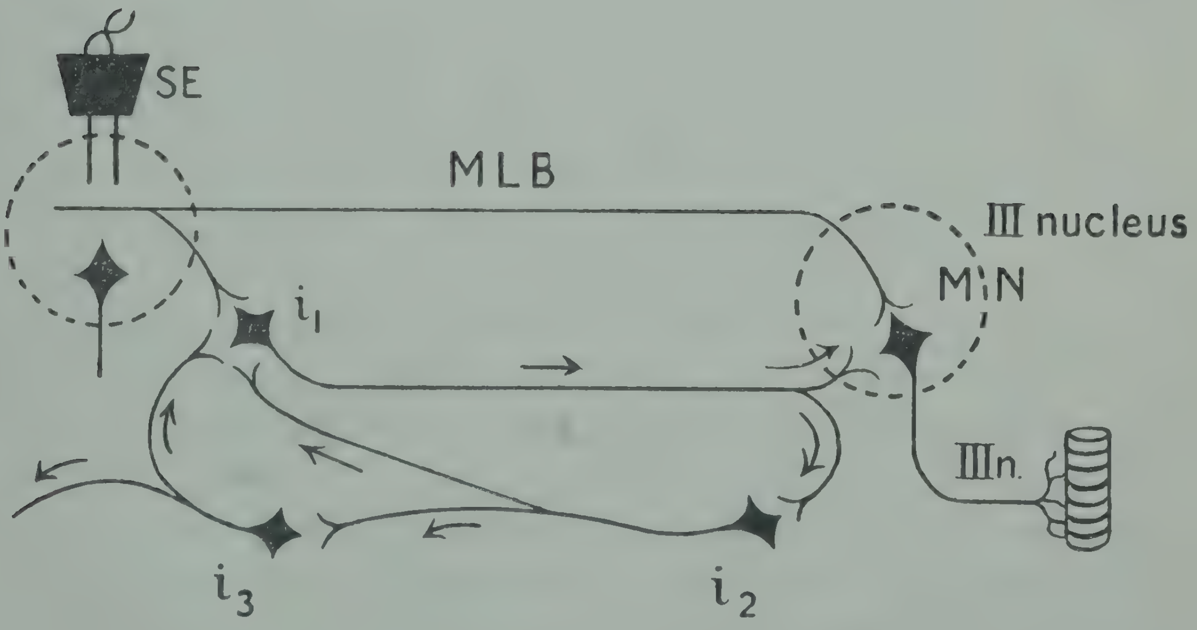

SE
MLB
III nucleus
i₁
MN
III n.
i₃
i₂

FIG. 41, 20

FIGS. 41, 18; 41, 19; 41, 20

Diagrams illustrating the experiments of Lorente de Nó. MLB, medial longitudinal bundle; SE, stimulating electrodes; MN, motor neurone; i_1, i_2, i_3, interneurones (much simplified from *Amer. J. Physiol.* 1935, **113**, 513; 1935, **111**, 276; *J. Neurophysiol.* 1939, **2**, 402; *J. cell. comp. Physiol.* 1935–36, **7**, 47).

of Fig. 41, 19. When a shock is sent in at SE an impulse travels directly to the motor neurone MN, raising its c.e.s. An impulse also travels to the interneurone i_1, which discharges after the synaptic delay and sends an impulse to MN. If a second impulse produced at SE reaches MN via MLB at the same time as an earlier one relayed in the interneurone, MN discharges and the medial rectus twitches.

Longer periods of facilitation can be explained if the internuncial neurones form closed self-re-exciting chains, as shown in Fig. 41, 20. Once such a chain is excited, it could send, via the branch of the axon of i_1, impulses to MN at intervals determined by the synaptic delays at i_1, i_2, and i_3, thus maintaining the c.e.s. for a considerable period. Since the impulses go round and round in the chains the process has been called *reverberation.*

CENTRAL SYNAPTIC TRANSMITTER

It is virtually certain that acetylcholine is the transmitter in the sympathetic ganglia (p. 748) and also at the neuromuscular junction (p. 788) but, although acetylcholine is present in the central nervous system, no clear demonstration that it acts as a transmitter at synapses within the brain or spinal cord has yet been given. Indirect but unsatisfactory evidence that it is a transmitter is given by the following experiments. When acetylcholine or an anticholinesterase such as eserine is applied to nerve cells, in the motor cortex for example, it depolarizes them, as it does muscle cells, and makes them discharge. Stimulation of the sciatic nerve of a dog whose spinal cord is perfused with eserinized saline increases the output of acetylcholine in the venous effluent.

Cholinergic nerves synthesize acetylcholine and can be identified by their high content of the enzyme choline acetylase. A high concentration is found in the anterior horns of the spinal gray matter and in the anterior roots of the spinal cord. This is in agreement with the fact that the anterior horn cells give rise to fibres which end at the neuromuscular junction where acetylcholine is actually liberated. The pyramidal tract containing the axons of the neurones from the motor cortex possesses little choline acetylase; the upper motor neurone is thus unlikely to be cholinergic. Intermediate values for the enzyme are obtained in the motor cortex which may be taken to mean that cholinergic nerves converge on the Betz cells. The afferent pathway is probably also made up of alternating cholinergic and non-cholinergic neurones. The first neurone from the periphery to the cord is non-cholinergic, the second is cholinergic, whereas the third, from the thalamus to the cortex, is again non-cholinergic.

Cholinergic nerves seem to contain acetylcholine all along their course and, correspondingly, adrenergic nerves contain noradrenaline. By analogy the transmitter substance of the primary sensory nerves might be expected to be present from the peripheral to the central end. Electrical stimulation of a sensory nerve sends impulses peripherally and releases a vasodilator substance resembling ATP. It has yet to be proved, however, that ATP is the transmitter released at the central end of the sensory neurones. The chemical nature of the non-cholinergic transmitter or transmitters in the central nervous system is still unknown.

Conditions influencing Conduction and Excitability

The speed of the passage of the nerve impulse is reduced when the temperature falls. A mammalian nerve cooled to between 0° to 5° cannot conduct impulses. Since the nerve recovers its excitability on being allowed to warm, this procedure is of value in experiments where a temporary arrest of nerve conduction is required, and refrigeration anæsthesia for surgical operations is based on this phenomenon. Another method of blocking nerve impulses is to narcotize the nerve by carbon dioxide, ether, chloroform, alcohol, cocaine, or procaine. All these substances have been used as anæsthetics and the two last are extensively employed to produce local anæsthesia so that minor operations may be carried out painlessly. Since efferent impulses are blocked at the same time, the muscles supplied by the nerve are paralysed and the blood vessels in the region become dilated.

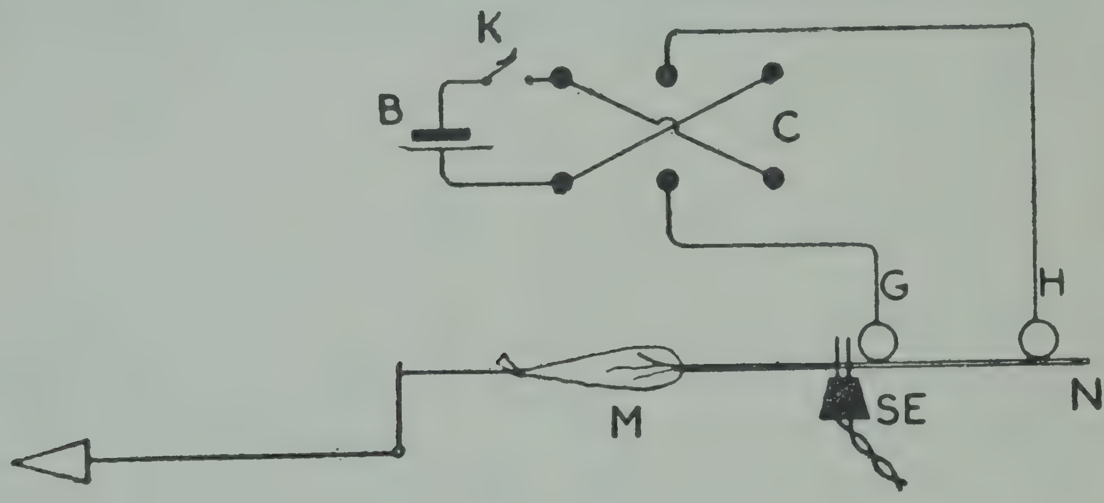


Fig. 41, 21

B, battery; C, reversing switch; G, H, nonpolarizable electrodes; M, muscle; N, nerve; SE, stimulating electrodes; K, key switch.

Conduction and excitability in a nerve can be modified by passing a constant electric current through it by nonpolarizable electrodes (Fig. 41, 21, G and H). A current applied in this way adds to or subtracts from the polarization of the resting nerve and so alters the conditions necessary to initiate a nerve impulse. Changes in the excitability may be studied by measuring the changes in the threshold to short shocks applied through another pair of electrodes SE. If G is negative the polarization of the membrane is reduced and a small current from an induction coil passed through the stimulating electrodes SE causes the muscle to contract. Conversely, when G is positive, a much stronger stimulus is necessary to cause contraction. The region of the nerve near the negative pole is said to be in a state of catelectrotonus, or increased excitability; the part near the positive pole has its excitability decreased and is in a condition of anelectrotonus. If G is made strongly positive and SE moved to a new position between G and H, or to the right of H, it is found that an impulse set up at SE is not able to cause a contraction of the muscle, that is the impulse is blocked by the region of anelectrotonus.

If the stimulating electrodes SE are removed from the nerve in

Fig. 41, 21, it is found that the muscle contracts when the circuit through BGH is completed by closing switch K and when it is opened by opening switch K; in other words, the muscle twitches at both 'make' and 'break' of the circuit. In both cases the impulse is initiated in the region of increased excitability, that is where there is a sudden appearance of catelectrotonus or a sudden disappearance of reduced excitability (anelectrotonus). The rise of catelectrotonus at make is a more effective stimulus than the fall of anelectrotonus at break because, if the current is weak, a contraction is obtained at make only and not at break. If, however, the current through the nerve is large when the electrode G, nearer the muscle, is positive there may be no contraction of the muscle at make; the impulse starting at the cathode H may not be able to pass the block due to the anelectrotonus at G. A contraction occurs at break because the impulse starting from the anode G at break meets no obstruction. When the current is strong with G as cathode the break excitation at the anode H may not pass the block at the area of depressed excitability at G due to the removal of catelectrotonus.

These results, first described by Pflüger in 1859, apply equally well to muscle and are easily verified on the frog's sartorius in which the fibres run from one end to the other of the muscle. If one end of the muscle is killed by heat, stimulation of this damaged area is ineffective. When a circuit is completed through the muscle with the cathode on living muscle and the anode on dead muscle a contraction is obtained. If the battery is reversed so that the cathode is on dead tissue there is no contraction at make. A make contraction thus starts at the cathode. Similar experiments at break show that a break contraction starts at the anode.

Strength-duration Relationship

In the experiment illustrated in Fig. 41, 22, an isolated nerve-muscle preparation is stimulated electrically through its nerve by using non-polarizable electrodes. The muscle M contracts only when the key K is closed or opened. With this arrangement the flow of current in the nerve begins abruptly when the key is closed and ends abruptly when the key is opened. When a steady current is flowing, that is when K is kept closed, the nerve is usually not excited and the muscle remains relaxed. In general terms, stimulation occurs only if the current is changing. If the key is closed and the potential across EE is made to increase very slowly from zero upwards by means of the potentiometer P, then the muscle does not contract until the voltage reached is much greater than that which is effective when applied suddenly. The nerve in some way adjusts or accommodates itself to the slowly changing current. Rapidly changing currents are thus more effective as stimuli than slowly changing ones.

By closing the key K at various positions of the contact on the potentiometer P we can find the smallest voltage necessary to stimulate the nerve and produce a minimal muscle twitch. This is the minimal or *threshold* stimulus. The threshold voltage for a current of infinitely long duration is called the *rheobase* or galvanic threshold. Currents of short duration are conveniently provided by releasing a pendulum (Fig. 41, 23), which in its swing knocks over two switches in succession to produce a current which rises abruptly to a level at which it remains constant until it ends abruptly by falling to zero. The duration of the stimulus is determined by the distance between the switches and its intensity by the voltage of the battery. The threshold voltage for short durations is higher than the rheobase, and the shorter the time of application of the current the higher must the voltage be to produce a minimal response. The relation is shown in the rectangular hyperbola of Fig. 41, 24, which is called a strength-duration curve. According to Lapicque the excitability of a tissue is indicated by its *chronaxie* which he defines as the shortest duration of the current of twice the rheobasic strength which is sufficient to excite the tissue. In Fig. 41, 24, the rheobasic strength is 2 units; since 4 units of current must be applied for 9 msec to produce a muscular contraction the chronaxie is 9 msec. The chronaxie is a convenient index of excitability since it is not greatly affected by the size of the electrodes or alterations in the resistance between the electrodes; also, the chronaxie of a muscle measured by stimulation through the skin is not very different from the value obtained by direct stimulation. The rheobase, on the other hand, is much more variable, since it depends very much on the experimental conditions.

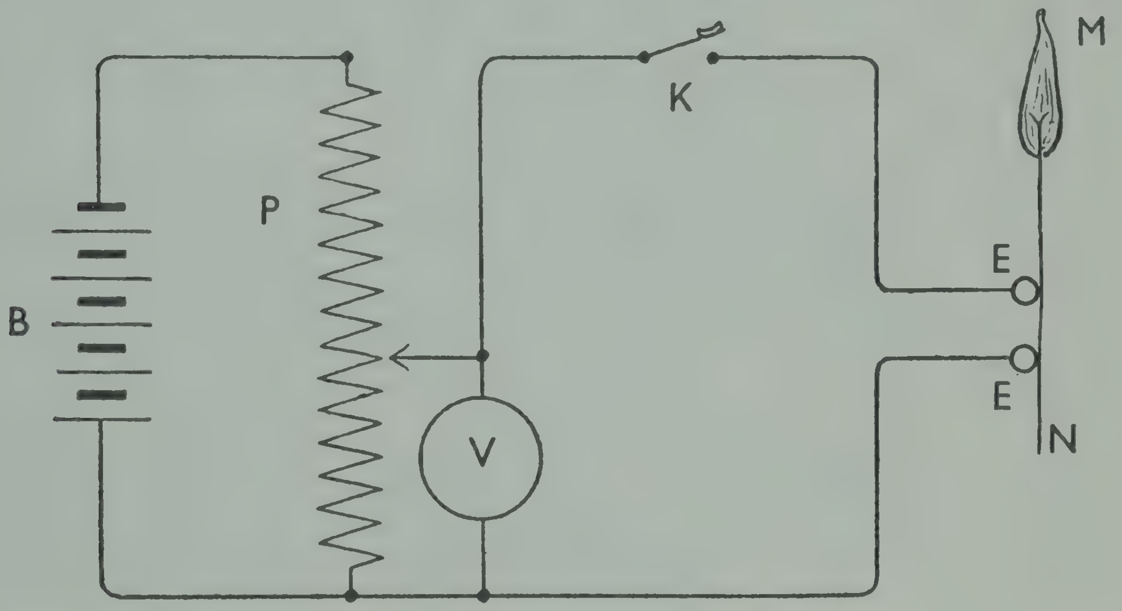


FIG. 41, 22

B, battery; P, potentiometer; V, voltmeter; K, key; EE, nonpolarizable electrodes; M, muscle; N, nerve.

The chronaxie is in general shorter the more rapid the action of a tissue, as shown in the following table:

Tissue	Chronaxie, approximate	Duration of Contraction, approximate
Frog's sciatic and gastrocnemius	0·5 msec	0·1 sec
Frog's ventricle . . .	3·5 msec	1·0 sec
Frog's stomach . . .	50·0 msec	15·0 sec

These experimental findings explain why it is sometimes reported that a tissue responds to galvanic currents but fails to react to faradic stimulation. The difference between these two currents is merely a matter of duration. A galvanic current is a steady current, usually taken directly from a battery. A faradic current is induced in the secondary winding of an induction coil by making or breaking the current in the primary winding and consists of one pulse or a short series of pulses of current each lasting about 0·1 msec. With so short a duration as this quite high voltages are needed to excite muscles and nerves but even high-voltage pulses from an induction coil cannot excite the slowly acting smooth muscle of the intestine.

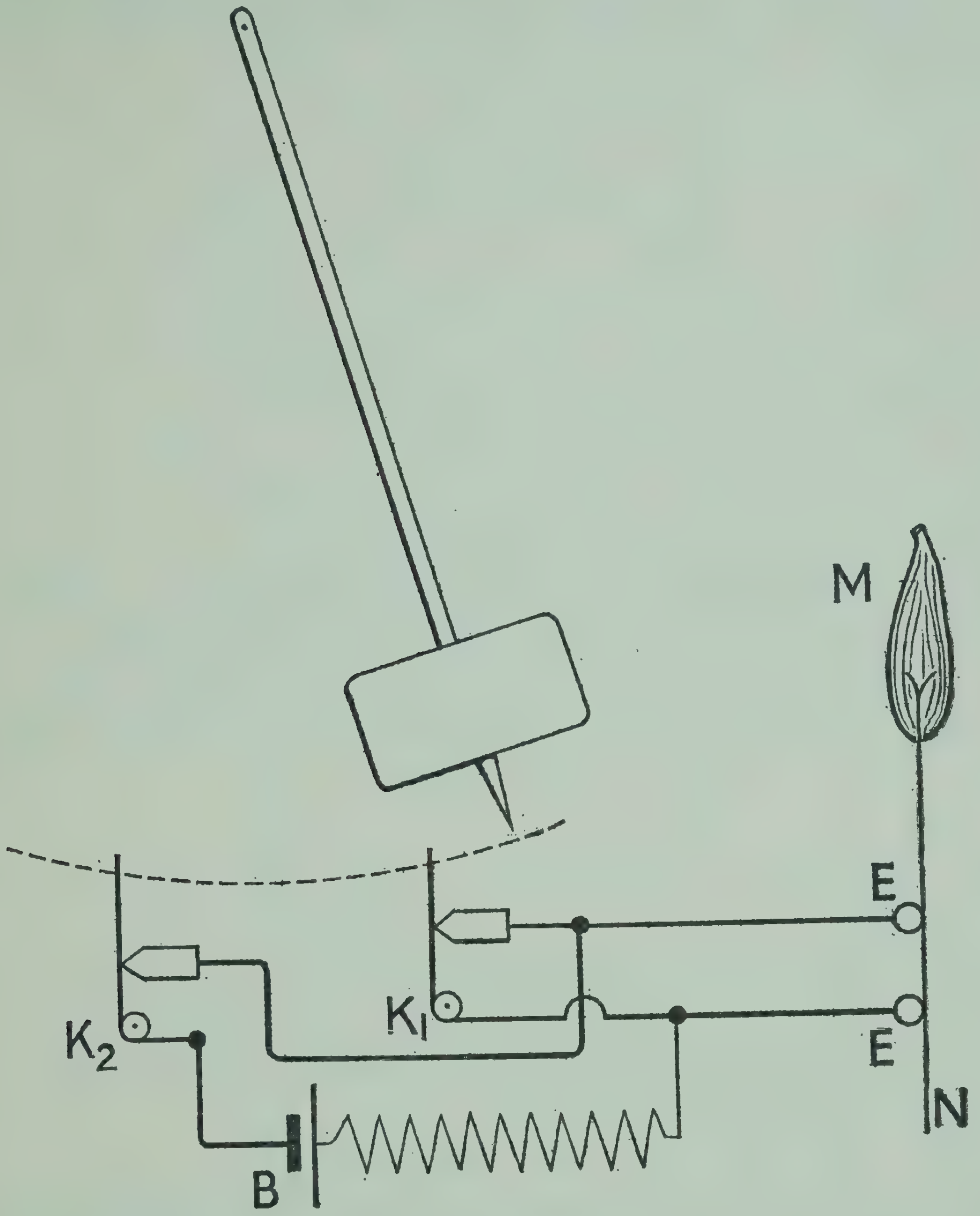


FIG. 41, 23

Scheme of apparatus for sending in brief rectangular pulses of current to a tissue. P is a heavy pendulum which on being released knocks open two switches, K_1 and K_2, in quick succession. When K_1 and K_2 are both closed the nerve N is short-circuited by K_1. When K_1 is knocked open current passes through the nonpolarizable electrodes EE into the nerve. When K_2 is knocked open the battery circuit is broken.

Very low-frequency alternating currents rise and fall too slowly to excite. High-frequency alternating currents fail to stimulate the tissues of the body because they do not last long enough in any one direction to reach the threshold. The most useful frequency for stimulation of mammalian nerves is 100 c/s. The main effect of radiofrequency currents is to heat the tissues through which they pass. Since the heating is followed by vasodilatation, high-frequency currents (diathermy) are useful in treatment, especially for deep-seated inflammatory conditions. If high-frequency currents are applied to the body through a needle electrode the heat generated locally can be made sufficient to burn the tissues and to produce coagulation of the blood and tissue proteins. Loss of blood at operation can be reduced to a minimum by using such a cautery instead of a scalpel.

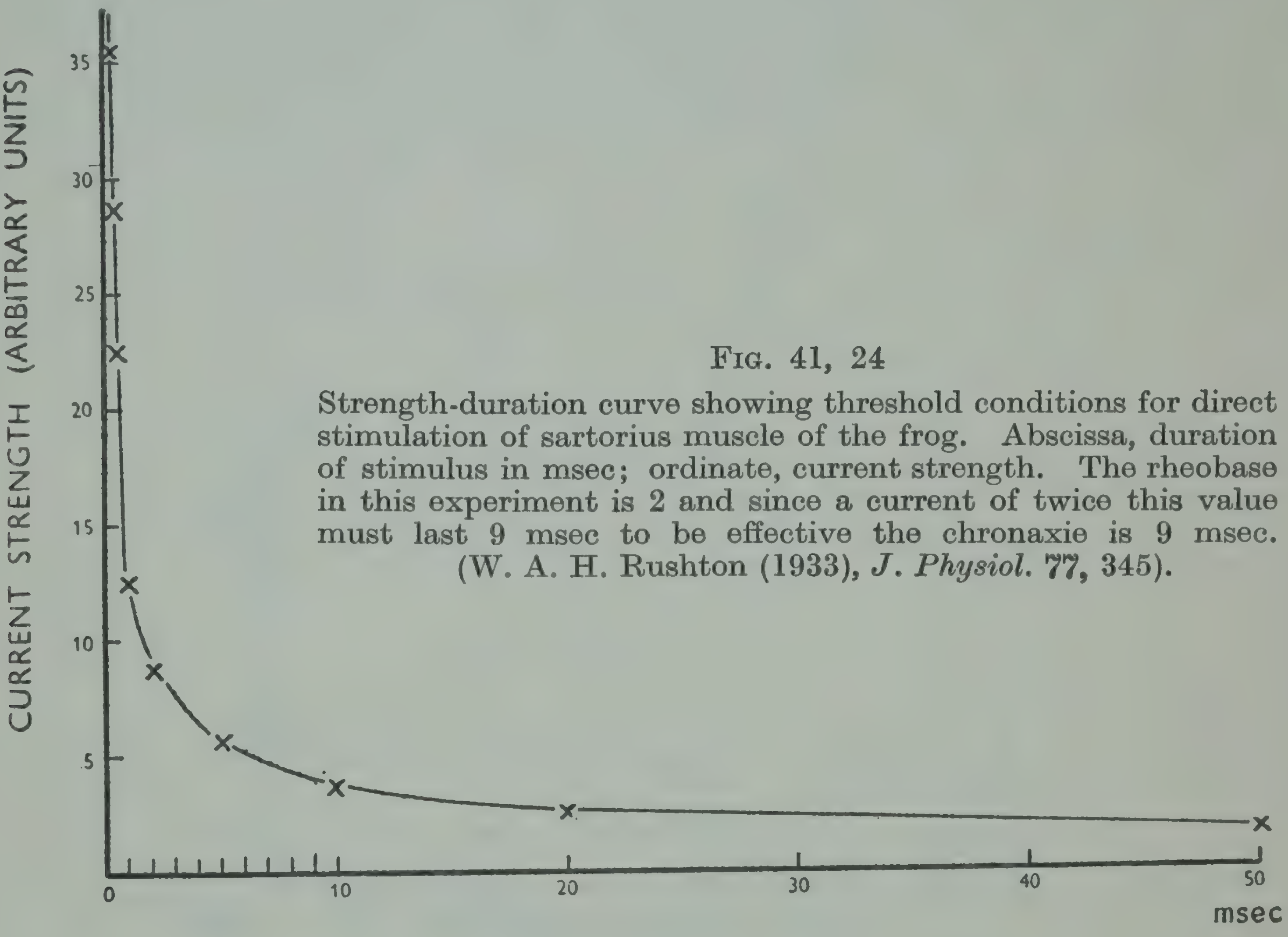


Fig. 41, 24

Strength-duration curve showing threshold conditions for direct stimulation of sartorius muscle of the frog. Abscissa, duration of stimulus in msec; ordinate, current strength. The rheobase in this experiment is 2 and since a current of twice this value must last 9 msec to be effective the chronaxie is 9 msec. (W. A. H. Rushton (1933), *J. Physiol.* **77**, 345).

It is important during the treatment of nerve injuries to study the effects of electrical stimulation of the nerves or muscles. In man the current is led into the tissues through electrodes placed on the skin. These are made of copper or brass discs covered with chamois leather which is soaked in saline. A small electrode to produce greater current density is placed over the nerve and a large electrode is placed at any other part of the body. The current may be considered to pass through the nerve as in the diagrammatic representation (Fig. 41, 25). It is found that excitation occurs with the weakest current when the electrode G is the cathode and the circuit is closed (CC). The order of increasing difficulty in eliciting a muscular contraction may be summarized thus:

CCC, or cathodal closing contraction.
ACC, or anodal closing contraction.
AOC, or anodal opening contraction.
COC, or cathodal opening contraction.

These results may appear at first sight to contradict the results of experiments on direct stimulation of nerve (p. 755) where it was shown that a make contraction starts at the cathode and a break contraction at the anode. In Fig. 41, 25, A, closing the circuit, CC, produces a region of catelectrotonus, that is a depolarization of the nerve, which sets off a propagated impulse. When the circuit is closed, AC as in Fig. 41, 25, B, a region of catelectrotonus is also produced but it is more diffuse; for a given flow of current the depolarization is less than in Fig. 41, 25, A and more current, or a higher voltage, is required to stimulate. Both CC and AC act by producing a depolarization at the place where the current leaves the nerve. Similarly it can be argued that AO is more effective than CO since with a given flow of current the anelectrotonus is more intense in the first case.

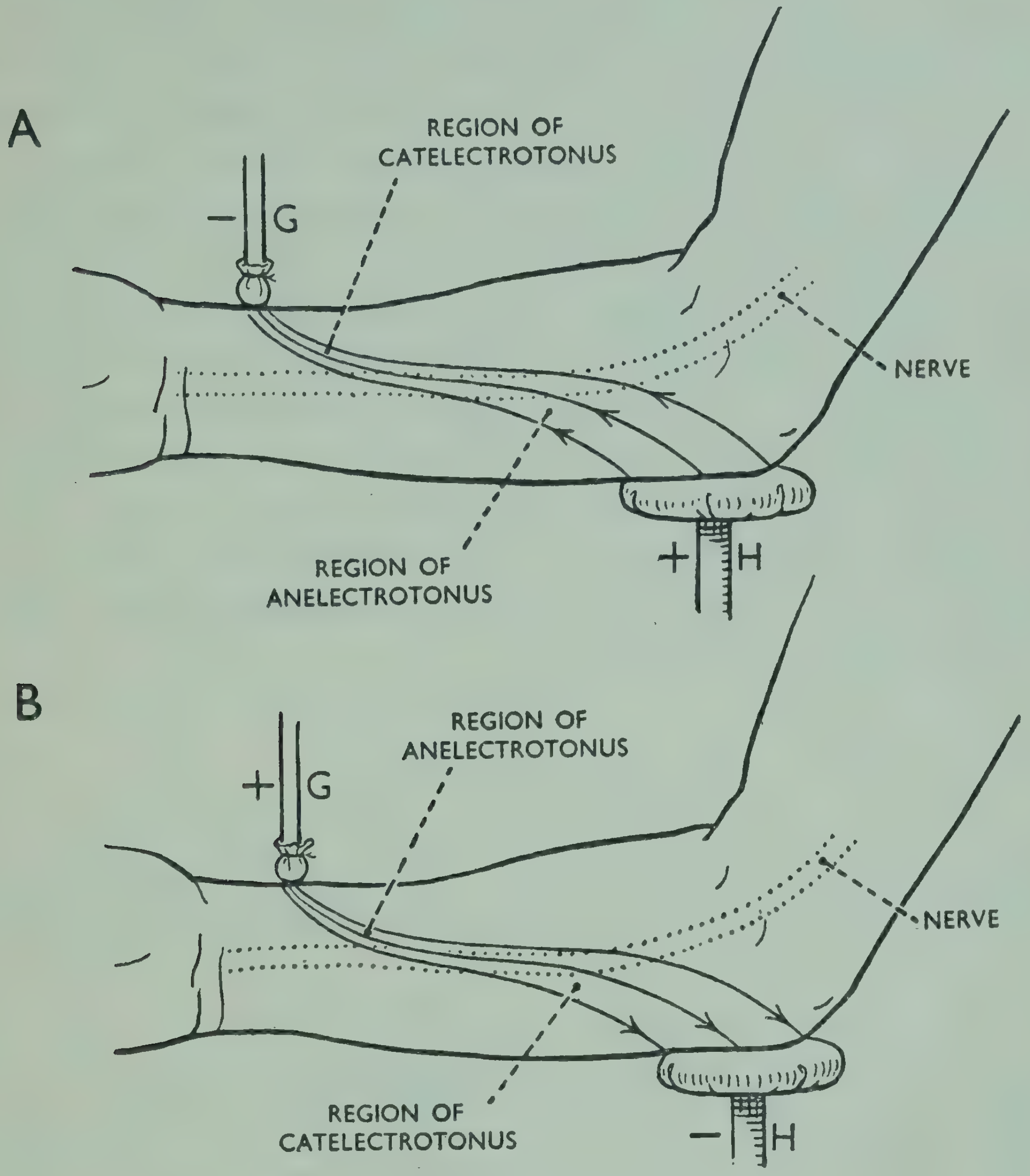


Fig. 41, 25

Diagram to illustrate the effects on a nerve of passing a current between chamois-covered electrodes G and H placed on the skin. The arrows show the conventional direction of the current, that is the direction of movement of a positively charged particle. A state of catelectrotonus exists where the current leaves the nerve; the region of catelectrotonus is more diffuse and, therefore, less intense in the lower diagram.

If now the electrodes, G and H of Fig. 41, 25, are connected to the secondary winding of an induction coil, it is found that the break 'shock' is more effective than the make 'shock.' This is not a contradiction of the statements of the previous paragraph which dealt with galvanic stimulation. It is due to the fact that when the flow of current in the primary winding is stopped by breaking the circuit the current induced in the secondary winding is very much greater than when it is made. The short pulses of current produced by an induction coil are effective during the rising phase of the pulse and we should regard excitation by induced currents as excitation occurring at make (CCC or ACC) as defined in the previous paragraph.

After degeneration of its motor nerve a skeletal muscle gives the *reaction of degeneration*, R.D. When surface electrodes are used as in

Fig. 41, 25 it is found that the muscle continues to respond to galvanic currents although faradic stimulation is ineffective (unless the voltage is extremely high). The chronaxie is increased perhaps 50 times and the strength-duration curve (Fig. 41, 24) becomes displaced to the right. Other signs of R.D. are a sluggish contraction and relaxation (occurring mainly under the stimulating electrode) in response to galvanic stimulation instead of the brisk jerk seen in normal muscle. The polar reactions just described may be altered so that ACC becomes equal to or greater than CCC but this is diagnostically unreliable. At the same time the denervated muscle becomes equally excitable at all points instead of showing maximal excitability at the point of entry of the motor nerve as does normal muscle.

The alteration in the strength-duration curve after denervation may be explained if it is assumed that the normal control muscle gives a curve characteristic of *nerve* whereas the denervated muscle gives a curve characteristic of denervated *muscle*. Although the curves do not, therefore, show a direct comparison between normal and denervated muscle yet the alteration in the strength-duration curve is of great diagnostic value. The rheobase of a denervated muscle stimulated *in situ* by external skin electrodes is usually higher than in a normal control but the rheobase of an isolated denervated muscle fibre is reduced presumably because of the increase in its membrane resistance.

Fatigue in Nerve

It is not possible in the presence of oxygen to fatigue a nerve by repeated stimulation; muscle on the other hand is very readily fatigued.

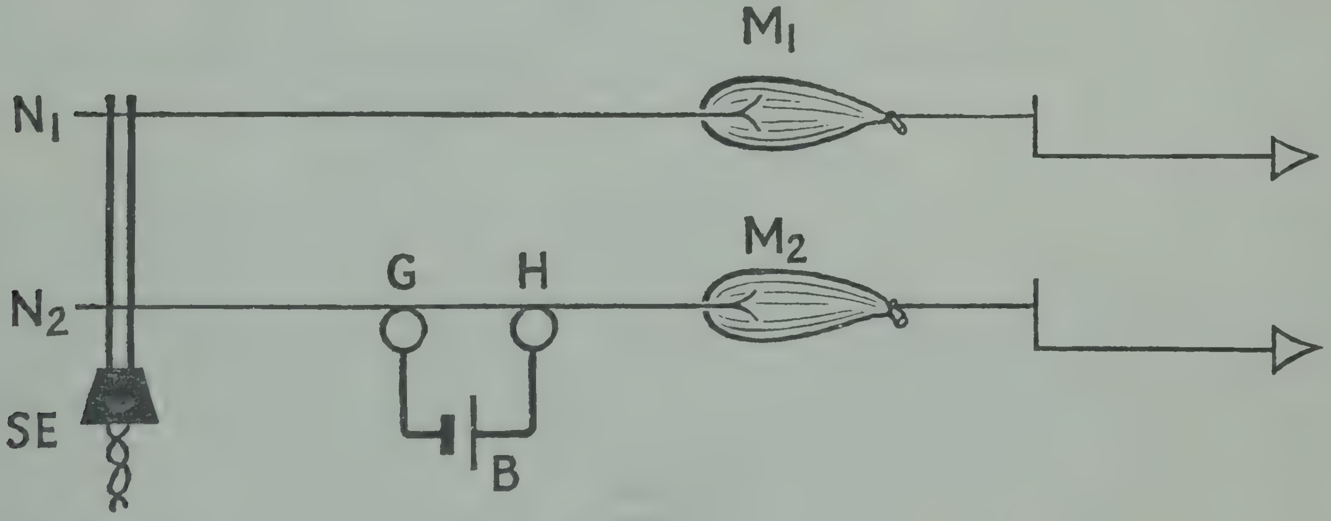


Fig. 41, 26

Experiment to demonstrate the site of fatigue in an isolated muscle-nerve preparation. N_1 and N_2, nerves supplying muscles, M_1 and M_2. SE, stimulating electrodes. G and H, nonpolarizable electrodes connected to battery B.

The difference between the two tissues can be shown by setting up two sciatic-gastrocnemius preparations as in the diagram (Fig. 41, 26). A series of stimuli sent through SE makes muscle M_1 contract, but muscle M_2 remains inactive because the nerve impulses set up in N_2 cannot pass the anelectrotonic block at H. After a time M_1 ceases to respond to stimulation at SE. If the battery B is now disconnected M_2 contracts vigorously, showing that the nerve N_2 between SE and H has not been fatigued although it has been carrying the same number of impulses as N_1. Since both muscles readily respond, if stimulated directly, the site of fatigue in this experiment must be the neuromuscular junction.

Further proof of this resistance to fatigue in nerves is shown by the fact that continued stimulation at fairly low frequencies does not lower the spike potential for a very long time. In the absence of oxygen, however, the response in such a preparation diminishes, showing that oxygen is required by the conducting mechanism. The reason for the slower onset of fatigue in nerves than in muscle is that the energy expenditure involved in a single nerve impulse is very much less than that expended in a single muscle twitch.

HEAT PRODUCTION AND METABOLISM OF NERVE

A. V. Hill and his colleagues, using specially designed thermocouples and galvanometers, showed that the conduction of nerve impulses was accompanied by the liberation of a minute amount of heat; it has also been shown that conduction of nerve impulses is accompanied by a small increase in the oxygen consumption and in the production of carbon dioxide. About one-twentieth (the exact fraction is rather variable) of the extra heat production occurs within one second and seems to be associated with conduction itself; this is called the *initial heat.* The remainder is liberated at a rate which rises quickly at first and then gradually falls off over the next thirty minutes; this is the *recovery heat.* The absolute quantities involved are exceedingly small, the resting heat production of a frog nerve at 20° C being about 70×10^{-6} calories per g. per sec, while the maximum additional heat production during continuous stimulation is about 40×10^{-6} calories per g. per sec. The total heat produced per impulse travelling in 1 g. of nerve is of the order of 10^{-6} calorie, which is equivalent to 10^{-12} calorie for a single impulse passing along 1 cm of a single fibre. The resting oxygen consumption of frog nerve is around 30 cu. mm per g. per hour and during maximal stimulation the consumption may be doubled. For frog muscle the resting heat production and oxygen consumption are of the same order but, since the amount of energy liberated by 1 g. of muscle in a single twitch may be many times that involved in the passage of a single nerve impulse, the rise in metabolic rate during activity may be much greater than it is in nerve. In the non-myelinated nerves of the crab, both the resting metabolic rate and the effect of stimulation on it are several times greater than they are in the frog nerve, for one of the advantages of the myelin sheath is that by restricting the flow of currents and the exchange of ions to the limited area of membrane at the nodes of Ranvier, it achieves an appreciable economy in the expenditure of energy. Mammalian nerves also have a more active metabolism than frog nerves but this is largely a consequence of the higher temperature at which they work.

After an impulse has passed along a nerve fibre it is left (p. 738) with slightly more sodium inside it, and slightly less potassium. This exchange of ions may be likened to the discharge of a chemical accumulator and constitutes the source of energy for the action potential. During the impulse there is probably both an absorption of heat (since the ions move into less concentrated solutions) and a liberation of heat (dissipated by the electric currents flowing in the neighbourhood of the active region), and the initial heat represents a balance between these in favour of the heat production. On the other hand the recovery heat, as its name implies, results from the operation of processes which restore the original condition of the axon—the most important of these probably being the extrusion of sodium and the reabsorption of potassium. Since both ions are being moved uphill, against their concentration gradients, the so-called 'sodium pump' requires a supply of metabolic energy which must ultimately be derived from the oxidation of foodstuffs. Very little is known in detail about any mechanism for the active transport of ions but, with the help of radioactive tracers, the rate of extrusion of sodium from stimulated squid and cuttlefish axons has been measured and it has been shown that the sodium pump can be brought virtually to a standstill by blocking metabolism with poisons like dinitrophenol and cyanide, and restored once more by washing the poison away. As metabolic inhibitors also greatly reduce the rate of entry of radioactive potassium, it seems that the extrusion of sodium is somehow coupled to an active uptake of potassium, both in nerve and in other types of cell.

In giant axons the recovery processes can be blocked by inhibitors without affecting the conduction mechanism. A poisoned axon can conduct hundreds of thousands of impulses with its sodium pump out of action because it has a large volume in relation to its surface area; the changes in the internal concentration for a single impulse are, therefore, very small. A further consequence of a large volume-surface ratio is that the rates both of running down during a bout of stimulation and of recovery after it are relatively slow; when the internal sodium is raised by a burst of impulses, it may be an hour or two before the original concentration is restored. The same is probably true for myelinated nerves where the restriction of ionic exchanges to the nodes results, in effect, in a large volume-surface ratio. In such nerves the time course of sodium extrusion is likely to be similar to that of the recovery heat. However, it is important to realize that these arguments do not apply in the case of nerve cells, which have a rather small volume and, because of their numerous dendrites, a relatively enormous surface area. Their internal ionic concentrations may be appreciably changed by the passage of only a few impulses, so that their recovery mechanisms must be constantly at work to keep them functioning and cannot be blocked for more than a few seconds without rendering them inexcitable. Other processes than ionic exchange are no doubt important in their working but these must be affected in a similar way by a small volume-surface ratio which is to be regarded as the primary reason for the marked dependence of the cells of the central nervous system on an adequate supply both of glucose and of oxygen. As will be appreciated from other chapters of this book, this dependence is the *raison d'être* for several vital regulatory physiological mechanisms.

The R.Q. of stimulated nerve is about 0·9, suggesting that at least part of the material being oxidized is carbohydrate, but the nature of the biochemical processes accompanying activity is still obscure. Although nerve tissue can produce lactic acid from carbohydrate in the absence of oxygen, the lactic acid so produced is oxidized only very slowly when access to oxygen is permitted, and there is little to suggest the presence of a glycogen-lactic acid system similar to that found in muscle. When a nerve is stimulated in the presence of oxygen, carbon dioxide, ammonia and inorganic phosphate are formed. It is probable that the phosphate is derived from the breakdown of ATP but the origin of the ammonia is not yet certain, although both adenylic acid and glutamine have been suggested as possible sources. Thiamine is probably necessary for the chemical equilibrium of the nerve cell and the effects of lack of thiamine have already been described on p. 122. Its mode of action is, however, still not clear.

REGENERATION OF PERIPHERAL NERVES

When a mixed peripheral nerve is cut across (*neurotmesis*) there is complete paralysis and anæsthesia in the area supplied by that nerve and the reflexes dependent on the nerve are abolished. The interruption of the efferent fibres causes paralysis of muscles, sudomotor, pilomotor and vasomotor paralysis; the interruption of the afferent fibres causes loss of tactile, pain, thermal, joint, postural and deep sensibility. Details of these types of paralysis and anæsthesia will be given later (p. 828). The peripheral part of the nerve, that is the part containing nerve fibres no longer connected to their cell bodies, degenerates, the affected muscles after a few days give the reaction of degeneration (p. 759) and begin to atrophy.

If a nerve is crushed, by the end of a fractured bone for example, the axons may be divided but the nerve trunk may be held together by its fibrous tissue. This lesion is described as *axonotmesis*. The paralysis and anæsthesia are exactly as in neurotmesis and, as before, the peripheral part of the nerve degenerates. The distinction is of practical importance because the effects of axonotmesis usually resolve spontaneously, whereas after neurotmesis a surgical operation is required to approximate the severed ends of the nerve.

Pressure on a nerve by a tourniquet or by a crutch may lead to the condition called *neuropraxia* in which there is predominantly motor paralysis with little muscular wasting and with relatively little sensory loss. The large myelinated nerve fibres are chiefly affected. The nerve does not degenerate and complete recovery usually occurs within a few days or at worst within a few weeks.

The nerve cell is remarkable in being anything from a thousand to a million times as long as it is broad. In spite of the great length of the cell, the cell body exerts a great influence over all parts, including the most distant. If a nerve is cut and examined about a day later it is found that material has been extruded from the end of the central stump. About twenty-four hours after section the myelin sheath in the peripheral stump breaks up into ovoid segments each smaller than the segments of Ranvier (p. 733). It has been suggested that the myelin is normally kept in long segments by the internal pressure and that after section of the nerve this pressure falls, surface tension then breaking up the fatty material into droplets. A week after section the degenerating myelin, after treatment with bichromate, stains black with osmic acid (Marchi's method). After a fortnight, when the fats are broken down still further, this staining property is gradually lost (see p. 824). Shortly after nerve section the axons also fragment and they and the myelin are gradually removed by macrophages. This whole process is often called *Wallerian degeneration* because it was described first by Waller in 1852. In the central stump, that part of the fibre attached to the body of the nerve cell, degeneration is usually confined to the last centimetre or so next the point of section, although a few of the central fibres and even their nerve cells may disappear. The fluid material of the axons (axoplasm) flows out from the central ends of the fibres and passes across the gap between the central and peripheral stumps into the tubes of neurilemma left after the degeneration of the peripheral stump. These tubes, which contain proliferated Schwann cells, guide the regenerating axons to the motor or sensory nerve-endings. The size of the new fibre is determined partly by the size of the parent fibre, partly by the size of the tube into which it grows, and also by the connexion made at the periphery, for if the fibre is able to function it becomes larger.

The growth of new fibres from the central stump down the peripheral scaffolding of Schwann tubes (or neurilemmal tubes) is apparently quite random. Thus sensory fibres may finally make connexion with muscle cells, in which case they cannot function and they are probably absorbed. If, when motor nerve fibres reach muscle cells, the end-plates have degenerated, new end-plates are formed; similarly sensory fibres growing to the skin may form new end-organs if the original ones have disappeared.

This randomness in the healing process excludes any regular order in the return of function and makes complete recovery most unlikely. Recovery of function after nerve section is, however, in many cases surprisingly good; this has been attributed to 're-education' of the nerve centres controlling muscles, although this notion is not generally accepted. The most complete restoration of function after nerve section is obtained when the cut ends are brought together very soon after the injury. The power of repair in nerves is not lost for many months but the longer the delay in apposing the ends the poorer is the final result. This is partly due to the atrophy of the end-organs and partly to the atrophy and eventual disappearance of muscle cells. Improvement in muscle power and in sensation often continues for up to three years.

Various methods are available clinically for determining the rate of recovery in a nerve. As regenerating fibres grow down the peripheral stump of a damaged nerve they re-innervate muscles in turn according to the distance to be travelled to each muscle. The time of recovery of motor function is taken as the time at which muscles show the first signs of voluntary contraction. This may precede the return of electrical excitability. From anatomical dissections detailed information has been obtained about the lengths of the nerves to the muscles and, by dividing the distance from the point of section to the point of attachment of the nerve to the muscle cells by the time needed for recovery, it is found that a nerve fibre grows about 1·5 mm per day. Percussion over the nonmyelinated tip of the advancing axon produces a tingling sensation which is known as Tinel's sign. This phenomenon advances at an average rate of about 1·7 mm per day, but it is not so satisfactory a method as the previous for determining the rate of regeneration in a nerve. In calculating the time likely to be taken after nerve suture for the return of function it is well to add six or eight weeks to allow for the possibility of an unusually slow rate of regeneration.

When the spinal cord is cut or the posterior roots are sectioned next to the cord, Wallerian degeneration occurs within the cord; although it may be followed by regeneration of a few new fibres within the central nervous system, restoration of function does not occur. The reason for this failure is not clear but it may be connected with the absence of neurilemmal cells from the central nervous system.

After section of a nerve fibre the cell body shows changes known as *chromatolysis* which is specially well marked if the section is near the cell body. The Nissl's granules disappear from the dendrites and later from the body of the cell itself and the nucleus moves to one side of the cell. These changes begin on the first day and reach their maximum in the third week, after which the Nissl substance gradually reappears and the cell regains its normal appearance.

The Afferent Impulse

So far we have considered the nerve impulse produced artificially by electrical stimulation of a nerve. In the living body certain nerve fibres are connected to specialized endings and these, when suitably stimulated, generate continuous trains or phasic discharges of afferent impulses. As Adrian pointed out, the stimulus applied to an end-organ may be compared to the pressure on the trigger of a machine-gun; either it is strong enough to fire the bullets or it is too weak to do anything at all (all-or-nothing relationship). Because of its refractory period the nerve fibre cannot transmit a steady potential but only a series of pulses, so that the message passing along the nerve may be likened to a stream of machine-gun bullets. If the abnormal conditions obtaining during the relatively refractory period are disregarded, the intensity of each impulse is constant or, to pursue the analogy, all the bullets are of the same size. The magnitude of the impulses cannot be altered by changing the strength of the stimulus. The only way, in fact, in which the message, carried by a given nerve fibre, can be graded to give the central nervous system information of the strength of the stimulus applied to the end-organ is by a variation in the total number of impulses and in the frequency with which they recur. Oscillograph records from even quite small nerves in the body may be too complex for analysis if many fibres are active and the action potentials from individual fibres are superimposed on one another. In the case of cutaneous sensory nerves the record may be simplified by stimulating only a very small area of skin, by cutting away unwanted receptors, or by cutting partly through the nerve. In this way the action potentials resulting from the activity of a single end-organ have been recorded. The form of these endings is very variable and, with the histological methods at present available, gives little clue to the particular stimulus to which the ending is sensitive. However, fibres have been isolated which carry signals from endings specifically sensitive to stretch, the chemical environment, temperature deviation, touch and pressure. The normal stimulus for the stretch sensitive endings is determined by the tissue in which they are deployed. Thus when arranged in the wall of a hollow organ, such as the carotid sinus, they signal fluid pressure (see Chapter 28, p. 474), whereas arranged on the fibres of a tendon they signal longitudinal tension. When linked mechanically to a contractile element composed of the intrafusal muscle fibres they act as muscle proprioceptors; a description of this type of ending, the muscle spindle, is given on p. 771.

Adrian and Zotterman elucidated many of the properties of the muscle spindle (p. 772) by using the small sternocutaneous muscle of the frog.

When the muscle was partly cut away records of the activity of one spindle were obtained from electrodes placed on the larger nerve trunk which is joined by the tiny nerve from the muscle. When the muscle was extended by a weight a regular series of nerve impulses was obtained. The frequency of the impulses was greater the greater the weight used (Fig. 41, 28), but it was always found that the frequency declined slowly from its initial high value (Fig. 41, 27) as shown by the greater spacing between the spikes

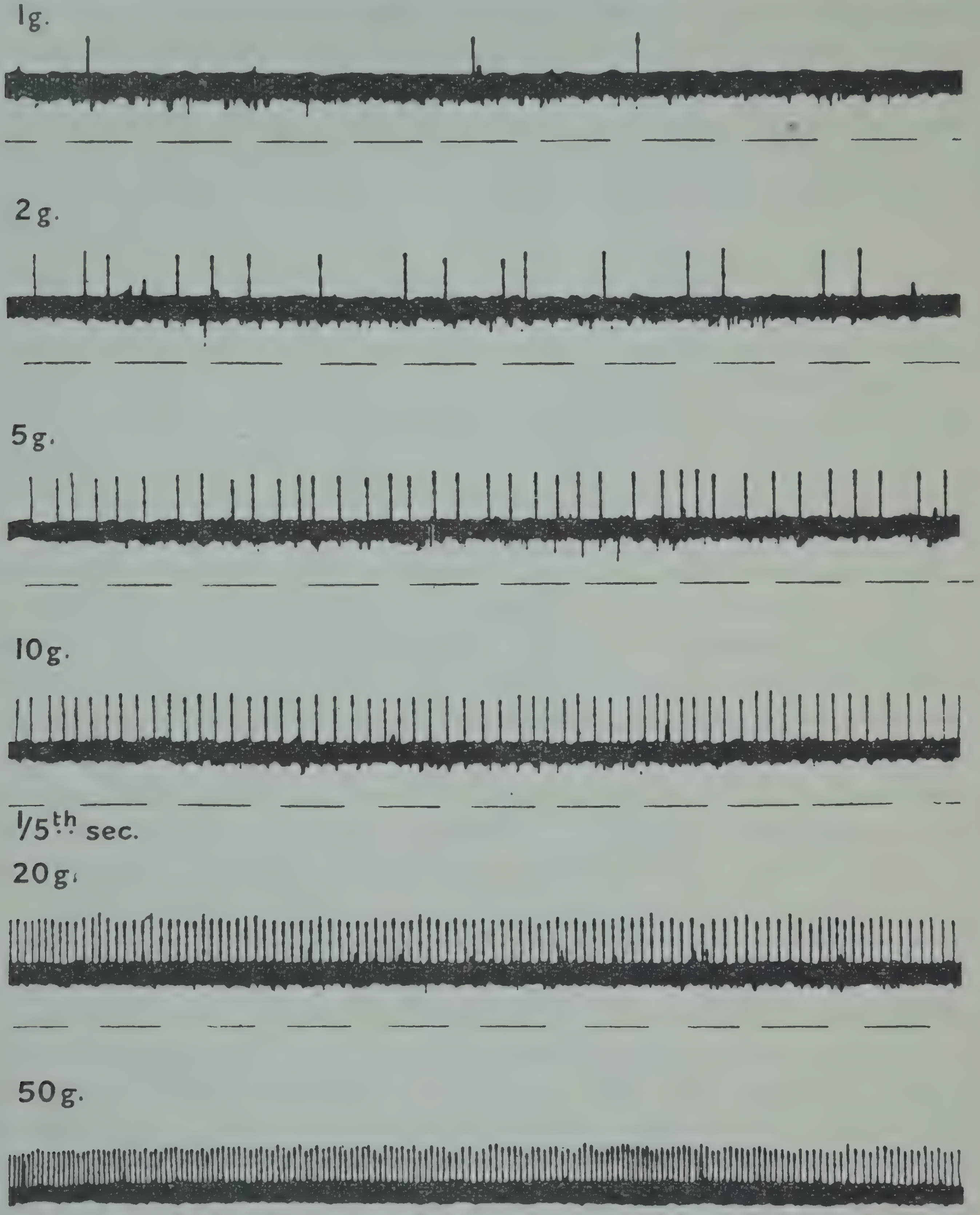


Fig. 41, 27

Photographic recordings from a cathode-ray oscilloscope of action potentials in the sciatic nerve from a single stretch-receptor in the gastrocnemius muscle of a frog. Responses in other stretch-receptors had been eliminated by cutting down the sciatic nerve trunk until it contained only one active fibre. The load indicated on the diagram was applied to the tendo calcaneus one second prior to the beginning of each record. Close inspection of the records, especially those for 20 g. and 50 g., shows that the space between individual spikes increases progressively from left to right. This decline in the frequency of discharge of the receptor is called adaptation. It is quite slow in the case of the stretch receptors. Some adaptation had, of course, occurred between the application of the weight and the beginning of the record in each case. Time marks represent $\frac{1}{5}$th sec., spike-height approximately 50 μV.

(*By courtesy of B. L. Andrew.*)

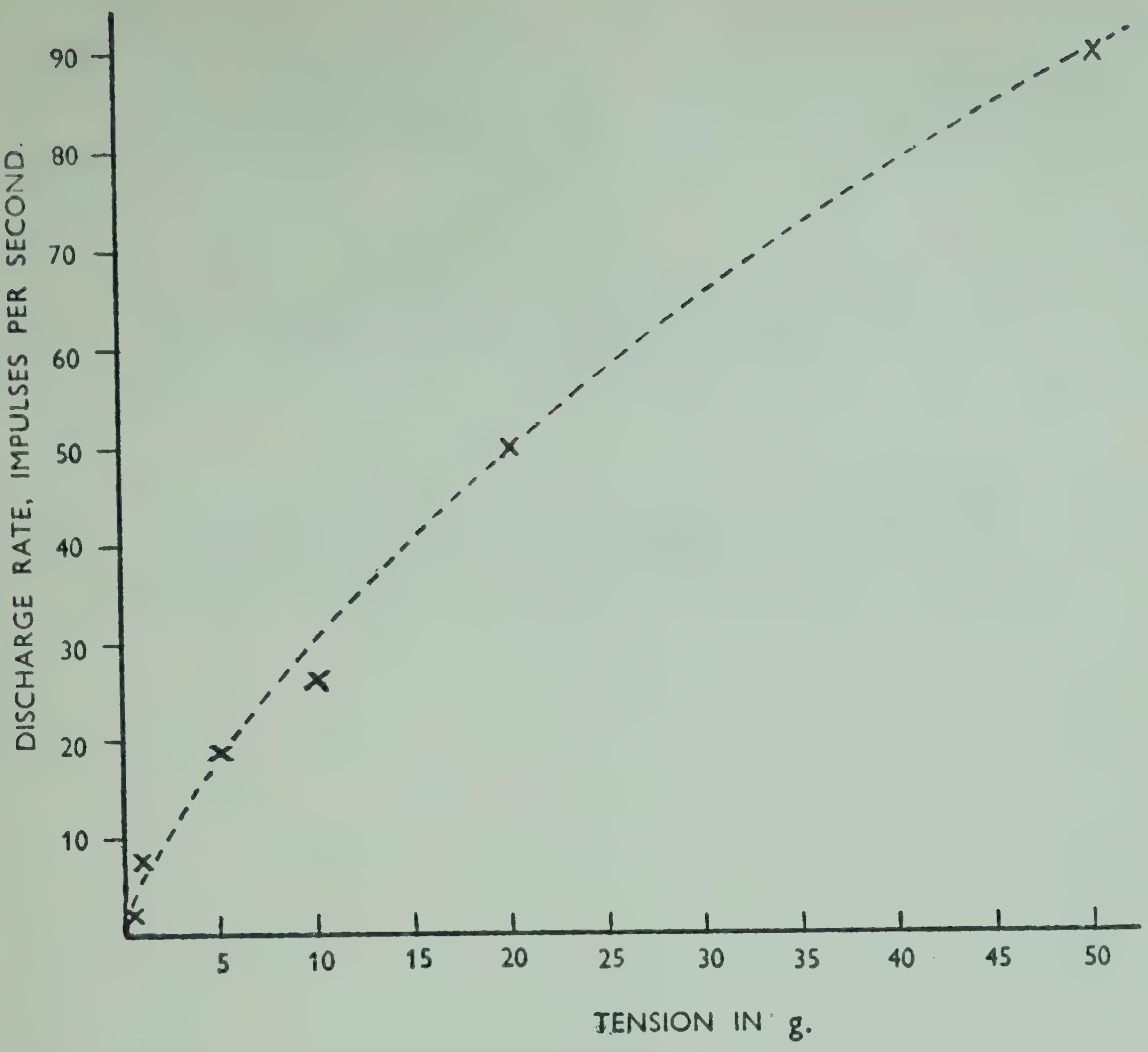


FIG. 41, 28

Graph showing the relation between the tension applied to the tendon of a frog's gastrocnemius and the rate of discharge of nerve impulses from a single stretch-receptor within the muscle (calculated from Fig. 41, 27).

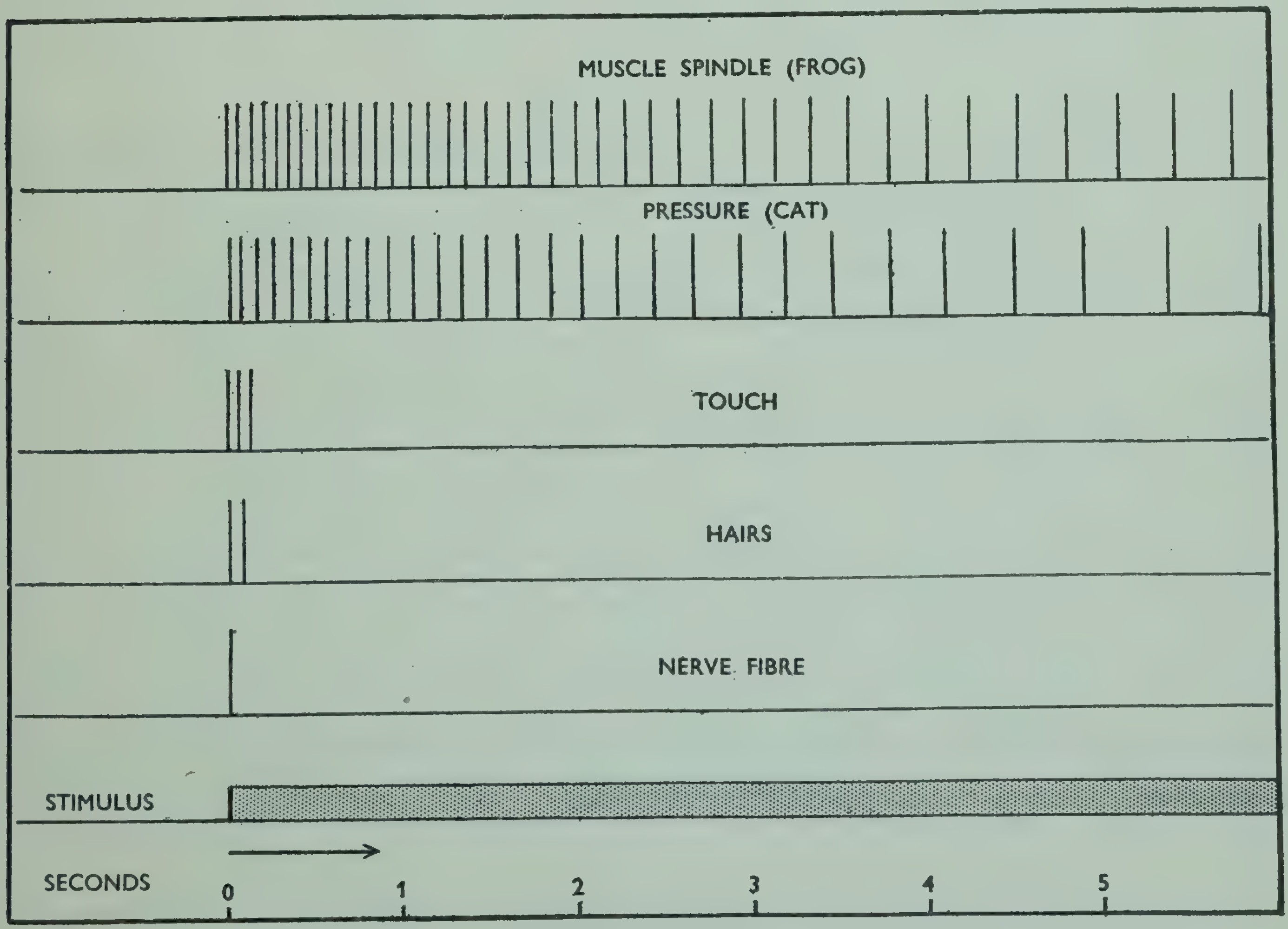

MUSCLE SPINDLE (FROG)
PRESSURE (CAT)
TOUCH
HAIRS
NERVE FIBRE
STIMULUS
SECONDS
0
1
2
3
4
5

FIG. 41, 29

Diagrams to illustrate the effect of applying a constant stimulus to various end-organs and to a nerve fibre. In the case of a nerve fibre a single impulse is aroused but in the end-organs a series of impulses of declining frequency is evoked. (*After* E. D. Adrian.)

towards the end of each record. This falling off in frequency of the sensory impulses is called *adaptation*; it is a decline in excitability produced by the stimulus but, considering what we know already about the physiology of nerve, it is almost certainly not due to fatigue. Each end-organ has its own peculiar rate of adaptation; muscle spindles adapt slowly, pressure organs more quickly and touch and hair endings very rapidly. The nerve fibre itself, in contrast with the nerve ending, adapts most quickly of all, since it responds with only one impulse at the beginning of a constant stimulus. These observations are summarized in Fig. 41, 29.

By placing electrodes near the muscle spindle in a frog muscle Katz has found that, when the muscle is stretched, there is in addition to the spikes (nerve impulses) shown in Fig. 41, 27 a negativity of the spindle. The amount of this electrical potential, which gives rise to a train of impulses in the sensory nerve, depends on the amplitude and rate of stretching. The spikes but not the local potential are abolished by a local anæsthetic. Katz has tried to account, at least partly, for these findings by supposing that the cell membrane becomes thinner on stretching so that its capacity increases and its potential (the local potential) falls. In the muscle spindle the stretch stimulus may be produced by activity of the contractile elements within the spindle or stretching of the spindle as a whole if the muscle is stretched; in the living animal both these processes may interact. A simpler arrangement is found at the joints where stretch receptors concerned with joint position sense are found on connective tissue elements such as the fibres of the capsule and the joint ligaments. The endings which, in any given position of the joint, lie on *stretched* connective tissue, discharge continuously. As parts of the joint are differentially stretched, or compressed, according to the position of the limbs, an individual joint proprioceptor may discharge impulses over a limited area of joint position (Fig. 41, 30).

An end-organ, such as a tactile end-organ, which adapts quickly cannot give well-graded responses to different strengths of stimulation because the response is over and done with in such a short time. On the other hand slowly adapting end-organs, like the muscle spindles, are able to give the central nervous system very accurate information about the intensity of the stimulus. Since a tactile stimulus is usually applied not to a single end-organ but to a fairly large area of skin a large number of end-organs is affected and so the information received at the central nervous system may be more complete than would be inferred from our knowledge of the properties of a single end-organ. Moreover when we touch an object we often run the fingers over it. This procedure causes stimulation of a large number of the quickly adapting end-organs and presents each with a rapidly changing stimulus so that each end-organ is continually re-excited. In this way a quickly adapting end-organ may be made to send a large number of impulses to the central nervous system.

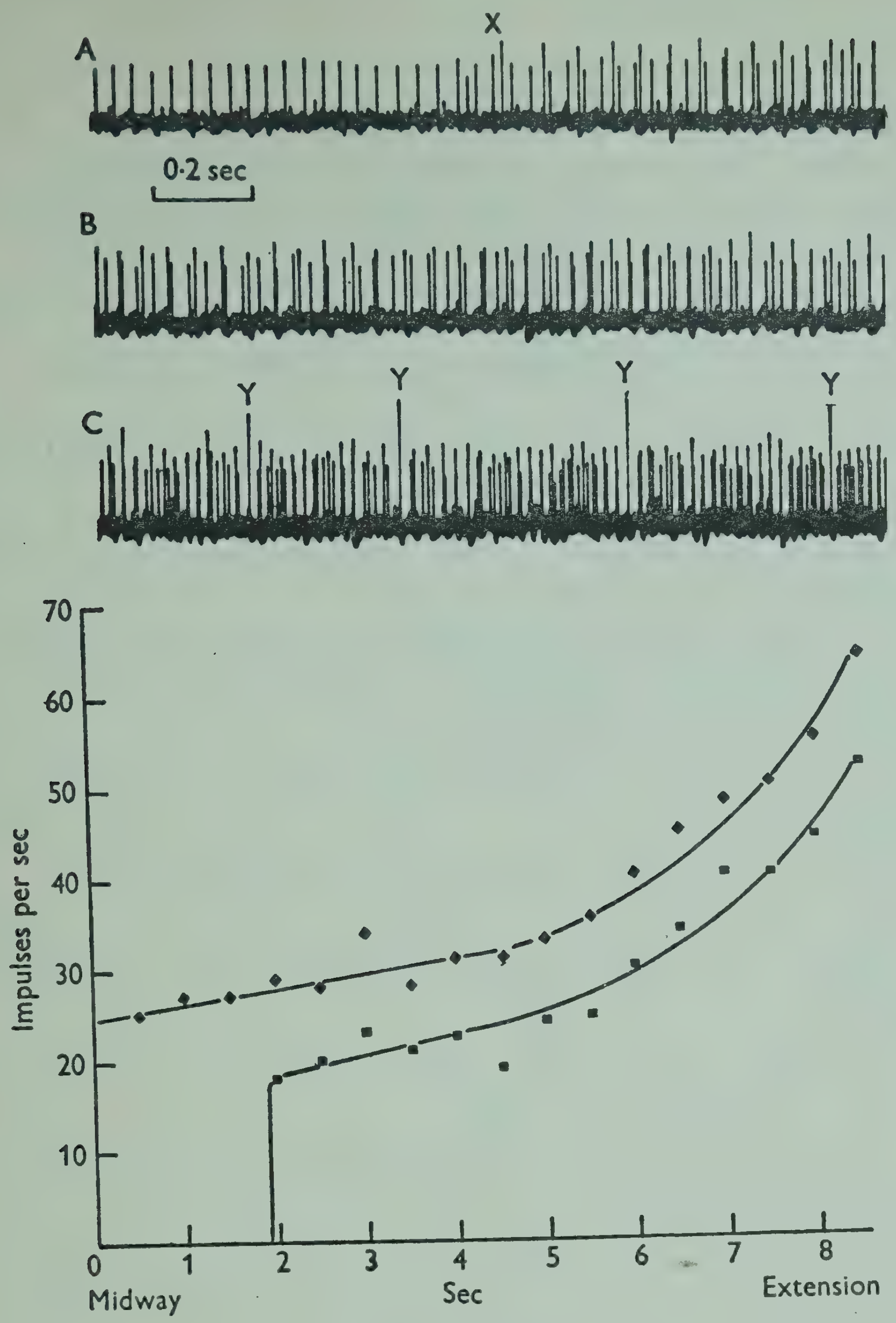


FIG. 41, 30

Records of action potentials taken from a small sensory nerve filament arising from sensory endings in the thyroepiglottal joint of the larynx of a rat. The joint was moved slowly to full extension, the whole movement lasting 8 sec. A, B and C are small sections of the original continuous record taken respectively at the beginning of the movement and 4 at 6·5 sec later. At the beginning of A, a single ending was active and discharged impulses at the rate of 25 per sec. Note that the impulses are of nearly constant size and are regularly spaced. At the point marked X a second ending began to discharge impulses which can be distinguished from those in the fibre connected to the first ending by the difference in the size of the action potential. The progressive addition of new active endings as the stimulus (in this case joint extension) is increased is referred to as recruitment of sensory endings. The large action potentials marked Y are not due to the recruitment of a third ending but to the electrical summation of two impulses which have occurred simultaneously in each of the two active fibres. In the lower figure the frequency of discharge of the two endings has been plotted against time. Apart from the differences of threshold the curves are of similar shape. As the frequency of both endings rises as the joint is extended it is easy to see that these endings signal the position of the joint to the central nervous system, that is they act as joint proprioceptors (B. L. Andrew (1954), *J. Physiol.* **126,** 514).

REFERENCES

Adrian, E. D. (1928). *Basis of Sensation.* London: Christophers.

Brazier, M. A. B. (1951). *The Electrical Activity of the Nervous System.* London: Pitman.

Brock, L. G., Coombs, J. S. & Eccles, J. C. (1952). The recording of potentials from motoneurones with an intracellular electrode. *J. Physiol.* **117,** 431–460.

Elliott, K. A. C., Page, I. H. & Quastel, J. H. (1955). *Neurochemistry: the Chemical Dynamics of Brain and Nerve.* Springfield: Thomas.

Erlanger, J. & Gasser, H. S. (1937). *Electrical Signs of Nervous Activity.* Oxford University Press.

Feldberg, W. (1945). Present views on the mode of action of acetylcholine in the central nervous system. *Physiol. Rev.* **25,** 596–642.

Feldberg, W. (1949–50). The role of acetylcholine in the central nervous system. *Brit. med. Bull.* **6,** 312–321.

Floyd, W. F. (1954). Electrochemical properties of nerve and muscle in *Modern Aspects of Electrochemistry.* Ed. J. O'M. Bockris (1954). London: Butterworth. Pp. 277–327.

Gasser, H. S. (1955). Properties of dorsal root unmedullated fibres on the two sides of the ganglion. *J. Gen. Physiol.* **38,** 709–728.

Granit, R. (1955). Receptors and sensory perception. London: Oxford Univ. Press.

Gray, J. A. B. & Sato, M. (1953). Properties of the receptor potential in Pacinian corpuscles. *J. Physiol.* **122,** 610–636.

Hill, A. V. *et al.* (1933). The three phases of nerve heat production. *Proc. Roy. Soc.* B, **113,** 345–356.

Hodgkin, A. L. (1949–50). Conduction of the nervous impulse. *Brit. med. Bull.* **6,** 322–325.

Hodgkin, A. L. (1951). The ionic basis of electrical activity in nerve and muscle. *Biol. Rev.* **26,** 339–409.

Hodgkin, A. L. & Keynes, R. D. (1954). Movements of cations during recovery in nerve. *Symposia Soc. exp. Biol., N.Y.* **8,** 423–437.

Katz, B. (1950). Action potentials from a sensory nerve ending. *J. Physiol.* **111,** 248–260.

Katz (1950). Depolarization of sensory terminals and the initiation of impulses in the muscle spindle. *J. Physiol.* **111,** 261–282.

Keynes, R. D. (1954). The ionic permeability of nerve membranes. *Colston Papers,* **7,** 43–49.

Kitching, J. A. (ed.) (1954). Recent developments in cell physiology. *Colston Research Soc. Symposium.* London: Butterworth.

Lorente de Nó, R. (1947). *A Study of Nerve Physiology.* New York: Rockefeller Institute.

McIlwain, H. (1955). *Biochemistry and the Central Nervous System.* London: Churchill.

Metabolism and function in nervous tissue (1952). *Biochem. Soc. Symp. No. 8.* Cambridge University Press.

Paton, W. D. M. (1952–3). The principles of ganglionic block. *Lect. Sci. Bas. Med.* **2,** 139–164.

Seddon, H. J. (ed.) (1954). Peripheral nerve injuries. *Spec. Rep. Ser. med. Res. Coun., Lond.* No. 282. London: H.M.S.O.

Stämpfli, R. (1954). Saltatory conduction in nerve. *Physiol. Rev.* **34,** 101–112.

The neuron (1952). *Cold Spring Harbor Symposia on Quantitative Biology,* **17,** 1–315.

Thomson, R. H. S. (1952–3). Cholinesterases and anticholinesterases. *Lect. Sci. Bas. Med.* **2,** 165–181.

Young, J. Z. (1942). Functional repair of nerve tissue. *Physiol. Rev.* **22,** 318–374.

Young, J. Z. (1949). Factors influencing the regeneration of nerves. *Adv. Surg.* **1,** 165–220.

CHAPTER 42

MUSCLE

STRUCTURE OF SKELETAL MUSCLE

EVEN the smallest skeletal muscle seen in an anatomical preparation is made up of very large numbers of muscle cells, frequently called fibres. The individual cells, which are from 1 to 40 mm long (occasionally 120 mm) and from 10 to 100μ in diameter, have a tough outer membrane, the sarcolemma, and each possesses numerous nuclei which lie immediately below this covering. The size of the cells and the amount of tissue lying between them, especially elastic fibres, both increase with age; a meshwork of connective tissue fibres between the contractile cells connects them to the tendon. The cell as seen in a histological preparation is filled with numerous sarcostyles (myofibrils) each approximately 1μ in diameter embedded in sarcoplasm. The sarcostyle itself consists of a concentrated gel of myosin while the sarcoplasm is a complex mixture of proteins including myosin. Striations run across the whole width of the cells breaking them up into isotropic I bands and anisotropic A bands. In ordinary light the A bands are dark and the I bands light but in polarized light the appearance is reversed. Each I band is crossed by a narrow but prominent Z band (Krause's membrane) while the A bands are similarly divided by M bands. All these show up clearly in an electron micrograph (Fig. 42, 1). Huxley has recently examined living frog muscle fibres with the interference microscope and finds that during passive stretch and in active contractions the A band remains approximately constant in length, the change in length occurring mainly by a change in length of the I band.

In many animals, the domestic fowl for example, red and white muscles can be clearly recognized. The red fibres have more muscle hæmoglobin and it has been found that red muscles contract more slowly and are less easily fatigued than white muscles. Although this differentiation is not strongly marked in human skeletal muscles it is likely that different parts of muscles are made up of cells with different rates of contraction.

Skeletal muscle cells are innervated by myelinated nerve fibres, 5 to 20μ in diameter, derived from the large motor cells of the anterior horn of the grey matter of the spinal cord. When a nerve fibre reaches the muscle its neurilemma merges with the sarcolemma under which the axon spreads out to form the motor end-plate (Fig. 42, 2). Each muscle cell may have one, two, or more end-plates; these may be supplied by branches of a single large nerve fibre or by branches derived from several nerve fibres (polyneuronal innervation). The blood vessels of the muscles receive efferent sympathetic nerve fibres but it is almost certain that the muscle cells receive no sympathetic supply.

In addition to the contractile fibres, muscles contain sensory elements called muscle spindles and Golgi tendon organs. Each mammalian *muscle spindle* is made up of several intrafusal muscle fibres enclosed in a fibrous tissue capsule. These fibres are of small diameter with some central nuclei and a collection of clear nuclei near the equatorial region of the spindle. The regions of clear nuclei are clothed with primary, annulo-spiral, nerve endings supplied by large nerve fibres (12 to 20μ); in addition, there are secondary nerve endings on the fine muscle fibres. The ends of the intrafusal muscle fibres are supplied with fine end-plates whose small motor nerve fibres (usually called the *gamma efferents* as they are about 5μ in diameter) leave the spinal cord by the anterior roots, independently of the large motor nerve fibres supplying the ordinary muscle fibres. The *tendon organs* are found near the musculo-tendinous junction and also in the fibrous tissue between the muscle fibres. They consist of irregularly arborizing nerve endings which penetrate between the tendon bundles. The annulo-spiral, secondary and tendon endings are all sensory since they degenerate after posterior root section. (The Figure 42, 2 shows the innervation of the muscle of the frog which differs slightly from that of the mammal as the intrafusal fibres are supplied by branches of the motor nerves to the ordinary muscle fibres.)

The muscle spindles lie parallel to the muscle fibres and so are altered in shape by any lengthening or shortening of the whole muscle. When a muscle is passively stretched the primary endings, which have a low threshold, signal the amount of stretching force by the frequency of the impulses sent up the sensory nerve. The secondary endings probably act in a similar way, but they have a higher threshold. When a muscle contracts it may shorten, the spindles are then relaxed and the sensory discharge stops. If, however, the gamma efferents are also stimulated, the sensory endings discharge again. This is presumably because the intrafusal muscle fibres contract, though no additional tension is recorded in the muscle; they take up the slack in the spindle and so restretch the sensory endings. In a muscle with an intact nerve supply any activity in the muscle has a reflex effect on the small nerve discharge, which in turn helps to grade the sensory discharges from the spindles. The discharges from the spindles help to excite the muscle in which they are situated and inhibit its antagonists (p. 845), while those from the tendon organs, which respond both to stretch and contraction of the muscle, inhibit the muscle and excite the antagonists. This reflex activity thus forms an integral part of the spinal reflex pattern.

There are also fine afferent nerve endings on the blood vessels which, one may speculate, may be stimulated by the products of the metabolism of muscle and may be concerned in carrying impulses to the brain which give rise to muscle pain.

Sherrington found that a quarter to one half of the fibres in a nerve supplying a muscle of a limb are sensory. Recently it has been discovered that only two-thirds of the motor fibres go to the muscle cells proper; the remaining third supply the intrafusal cells of the muscle spindles to adjust the sensitivity of these sensory organs to the needs of the movement being carried out. Thus more than half of the nerve supply of a skeletal muscle is concerned with the regulation of the discharge of the anterior horn cells.

CHEMISTRY OF SKELETAL MUSCLE

Skeletal muscle contains about 75 per cent. of water, 20 per cent. of protein, and 5 per cent. of other materials. When fresh muscle, finely minced with ice and a little sodium chloride, is put under pressure, a viscous

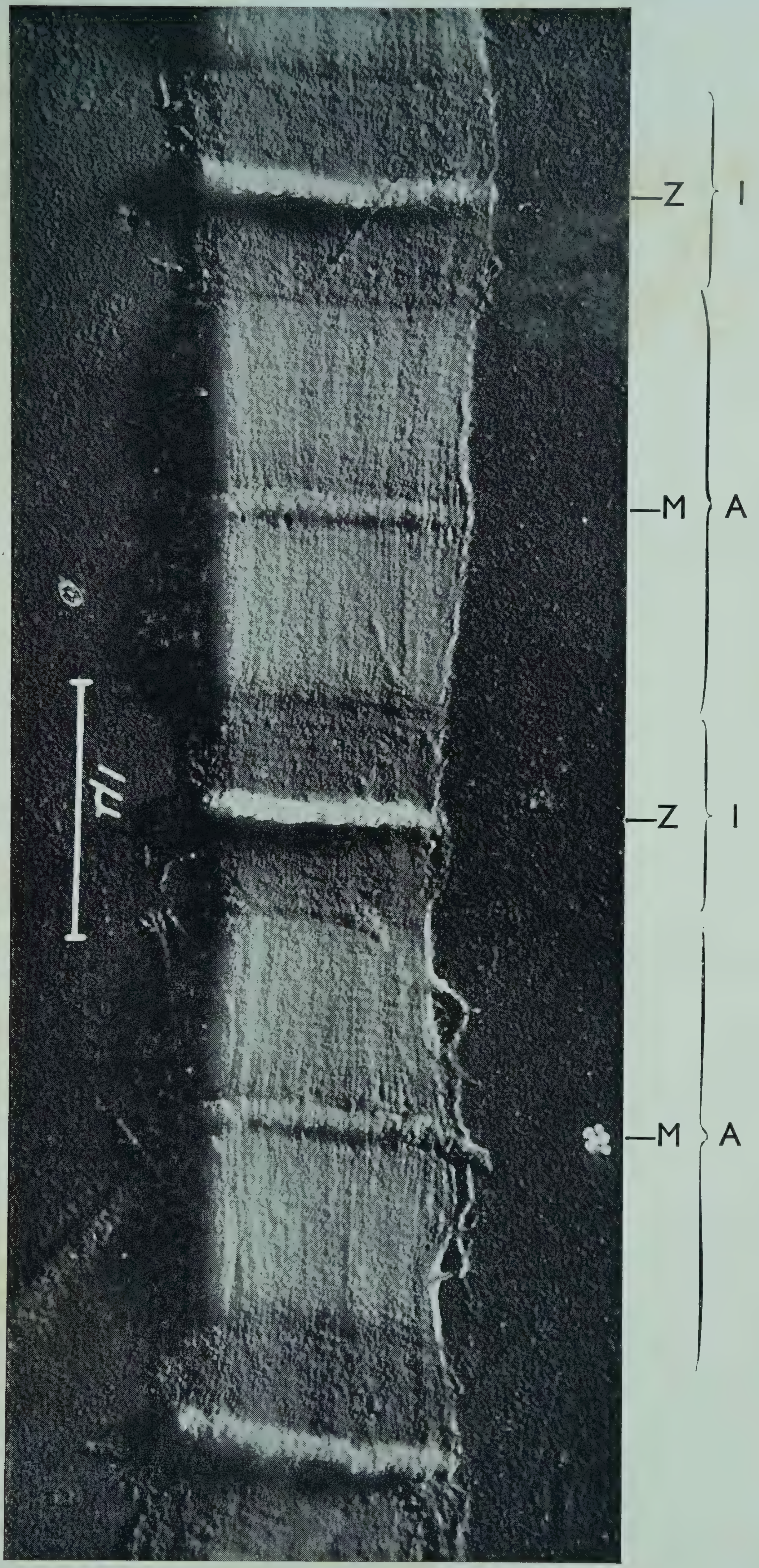


Fig. 42, 1

Electron micrograph of toad muscle, platinum shadowed. I, isotropic band. A, anisotropic band. Z, Krause's membrane. M, M-band.

(*By courtesy of M. H. Draper and A. J. Hodge.*)

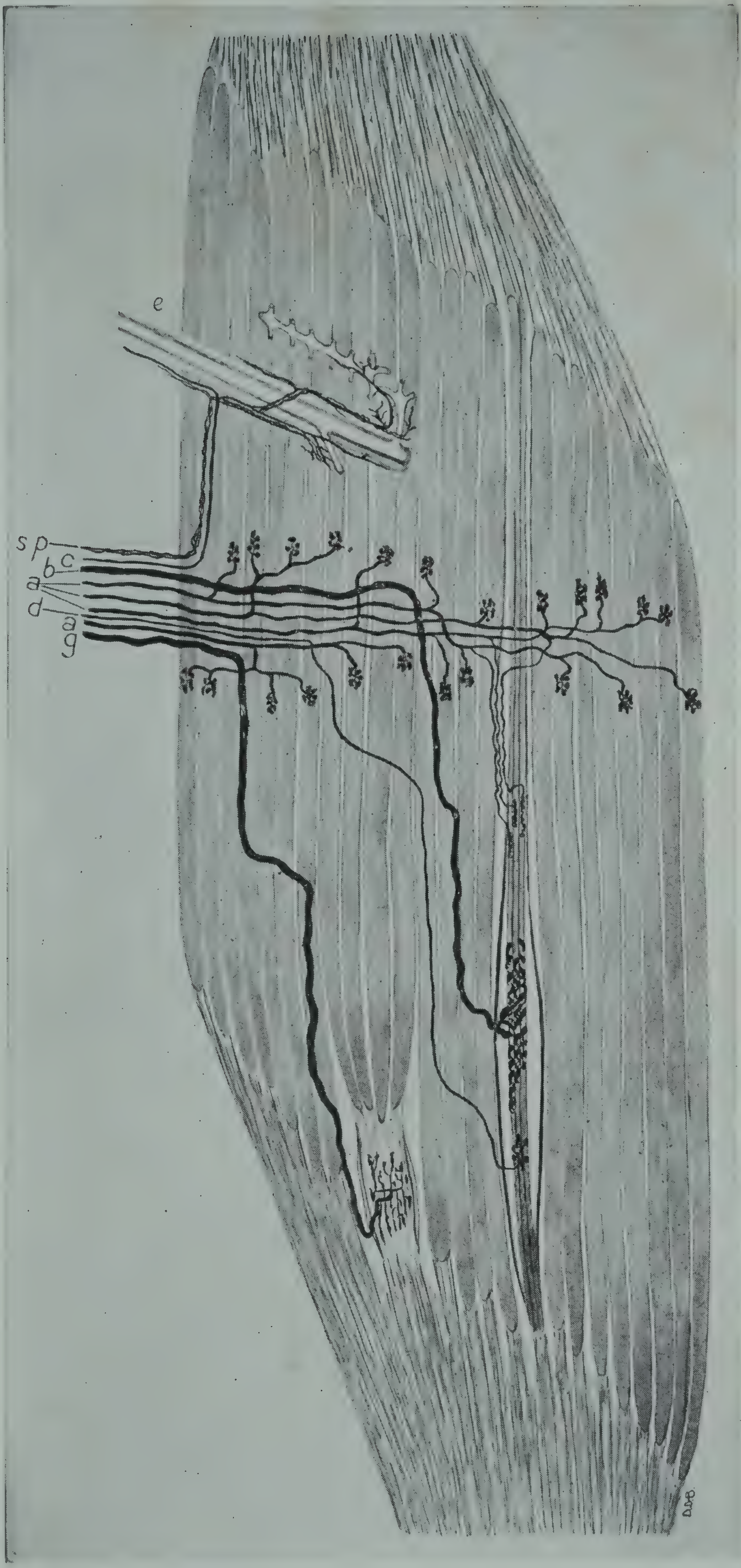


FIG. 42, 2

Diagram illustrating the innervation of a group of frog muscle cells. The motor innervation comes from fibres *a*, *a*, *a*, *a*; the muscle spindle is made up of 3 intrafusal fibres, each one innervated by branches of motor nerve fibres *a*, *a*, *a*, *a*. The spindle has one large annulospiral ending, *b*, and a single flower-spray ending connected with the nervous sytem by a fibre of relatively small diameter, *d*. There is a single Golgi tendon organ, *g*, and a sympathetic plexus, *sp*, accompanied by a small myelinated pain fibre, *c*, from the blood vessel, *e* (Creed, Denny-Brown, Eccles, Liddell and Sherrington (1932), *Reflex Activity of the Spinal Cord*, Fig. 71, facing p. 166. Oxford: Clarendon Press. In the mammal the intrafusal fibres are supplied by small motor fibres about 5 μ in diameter arising in the spinal cord.

solution of the proteins can be expressed ('muscle plasma'). This solution forms a gel if allowed to stand for a short time at room temperature.

A large proportion of the muscle protein consists of *myosin*, a globulin which is part of the contractile element itself. It is insoluble in water but it dissolves in dilute salt solutions, giving a viscous solution which when heated or diluted forms a gel. Solutions of myosin exhibit the phenomenon of streaming birefringence which is, however, reduced on addition of salts of adenosinetriphosphate (ATP). Myosin is made up of polypeptide chains lying along the axis of the sarcostyle which are normally folded in the α form but can be extended on stretching to the β form (*cf.* keratin, p. 48). During contraction of the sarcostyles, the polypeptide chains of myosin contract to a length still shorter than the α form (Fig. 42, 3).

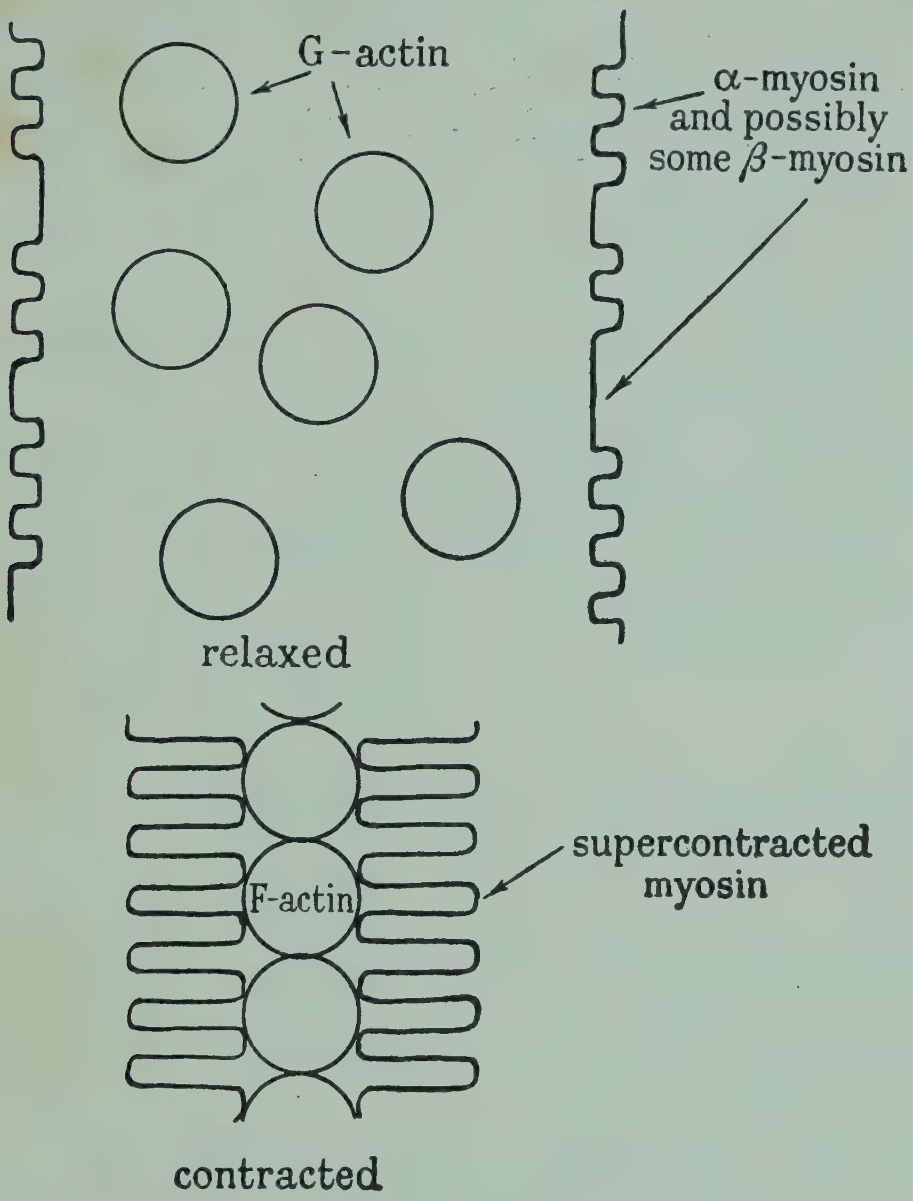


FIG. 42, 3

Diagrammatic representation of a possible mechanism of muscular contraction by the formation of an actomyosin complex (W. T. Astbury (1950), *Proc. Roy. Soc.* B, **137**, 61).

Szent-Györgyi and his associates have drawn attention to the importance in muscle of the protein *actin* which forms together with myosin a complex known as *actomyosin*, the actual contractile element. Actin itself exists in two forms, globular or G-actin and fibrous or F-actin. The latter, which has a much higher molecular weight than the former, is the form which combines with myosin.

Myosin and actin together comprise some 40 per cent. of the total muscle protein. Of the remainder 20 per cent. is *albumin*, 20 per cent. is *globulin X*, 20 per cent. is stroma protein (from the sarcolemma and connective tissue) and 6 per cent. is *tropomyosin*. All the enzymes which catalyse the various steps in the process of glycolysis are included in the protein fraction.

Muscle also contains in the sarcoplasm of the cell the respiratory pigments *myohæmoglobin* (*myoglobin*) and *cytochrome* which function in the transport of oxygen from the blood to the oxidizing systems.

After certain forms of injury (automobile accidents, air-raid casualties) there may be prolonged crushing of muscle, which reduces its blood supply and finally brings about the death of the tissue. In such circumstances myohæmoglobin diffuses into the general circulation and is excreted in the urine. A rather similar condition occurs in horses given hard exercise after a period of rest and rich feeding.

Only small amounts of lipid are found in muscle. Carbohydrate is present mostly as glycogen in amounts between 0·5 and 1·0 per cent.

Many of the simpler non-protein constituents of muscle are soluble in water and are usually termed *extractives*. They include *creatine* and *phosphocreatine* (*phosphagen*) (p. 352), *adenylic acid* and *adenosine triphosphate* (p. 60), *lactic acid*, *inositol*, *carnosine* and *anserine*.

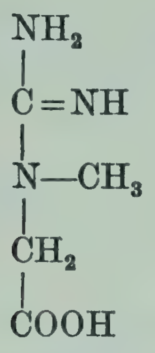
One of the most important compounds in muscle is adenosine triphosphate (ATP) in which the terminal phosphate group is linked by an 'energy-rich bond' (see pp. 60 and 318). The rupture of this bond with the liberation of inorganic phosphate and the formation of adenosine diphosphate (ADP) is associated with the release of a large amount of energy (11,500 cal per mole). It was shown by Engelhardt and Lyubimova in 1939 that myosin is associated with the enzyme adenosine triphosphatase which catalyses the breakdown of adenosine triphosphate (ATP) to adenosine diphosphate (ADP) and inorganic phosphate with the liberation of a great deal of energy:

$$\text{ATP} \longrightarrow \text{ADP} + H_3PO_4 + 11{,}500 \text{ cal.}$$

It is this energy released by the breakdown of ATP that is used in the performance of muscular work. By a method similar to that employed in the manufacture of artificial silk fibres, Engelhardt and Lyubimova prepared artificial myosin threads containing adenosine triphosphatase by forcing a myosin solution through a jet into an appropriate gelling medium. Small weights were then attached to the myosin fibres. When adenosine triphosphate (ATP) was added to the solution in which they were suspended, the fibres relaxed and lengthened. Other compounds, including adenosine diphosphate (ADP), had no effect. More recent experiments by Szent-Györgyi with actomyosin threads have made it clear that, when ATP is added to a thread to which no weight is attached, the thread contracts. The same effect is found with intact myofibrils. The contraction, which is accompanied by the hydrolysis of ATP, occurs only when magnesium and potassium ions are present in concentrations similar to those found in muscle cell fluid and is inhibited by calcium ions.

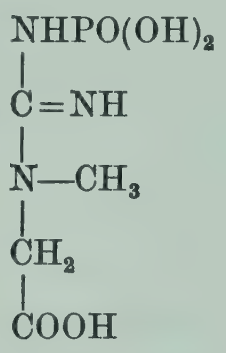
This sensitivity to changes in the concentration of ions in the environment, particularly potassium and calcium ions, is characteristic of muscle proteins. Although our knowledge of this subject is far from complete it is probable that the nerve impulse by bringing about some small local changes in the concentration of inorganic ions is able to operate the trigger mechanism which results in the contraction of the fibrils. Astbury has suggested that when muscle fibrils contract G-actin becomes aggregated to form fibres of F-actin. At the same time the myosin chains become supercontracted and attached to the F-actin threads (Fig. 42, 3).

It has been mentioned earlier that muscle tissue contains a large amount of creatine, most of which is present as *phosphocreatine* (*creatine phosphate* or *phosphagen*).



$$\begin{array}{l} NH_2 \\ | \\ C{=}NH \\ | \\ N{-}CH_3 \\ | \\ CH_2 \\ | \\ COOH \end{array}$$

creatine



$$\begin{array}{l} NHPO(OH)_2 \\ | \\ C{=}NH \\ | \\ N{-}CH_3 \\ | \\ CH_2 \\ | \\ COOH \end{array}$$

phosphocreatine (phosphagen)

During muscular contraction phosphocreatine breaks down to creatine and phosphoric acid. On recovery, phosphocreatine is resynthesized. There is evidence that phosphocreatine acts as a reserve or emergency source of phosphate for the synthesis of adenosinetriphosphate, a sufficient supply of which is vitally important:

phosphocreatine + adenosinediphosphate (ADP) ⟶ creatine + adenosinetriphosphate (ATP).

It was shown by Lundsgaard that, in muscles poisoned by sodium iodoacetate, lactic acid is not formed on contraction. In such cases energy can apparently be derived from the breakdown of phosphocreatine, which therefore provides an alternative source of energy in an emergency.

In invertebrates creatine phosphate is replaced by arginine phosphate.

The considerations described above allow the following tentative summary of the mechanism of muscle contraction. The myosin or actomyosin fibrils are normally in an extended form, but on the arrival of the nerve impulse they contract and perform work or, if contraction is impossible, they develop tension. The contraction energy is of course derived from the breakdown of ATP. It should be noted here that biochemical opinion is tending to the view that the neutral position for the muscle fibril is the 'contracted' state and that metabolic energy is required rather to return the elastic fibrils back to the original stretched or 'relaxed' state than to bring about the shortening or contraction of muscle. In other words, the resting muscle resembles a stretched elastic band and the external force which the contracting muscle exerts results from the recoil of its stretched fibrils. On this view 'relaxation' is an active process involving the elongation of the myosin molecules with the aid of the energy of the ATP. This, however, is difficult to reconcile with A. V. Hill's findings described on p. 792.

Whatever view is finally adopted to account for the initial step, it is clear that the breakdown of ATP is the fundamental chemical process which supplies the energy. Subsequent steps are concerned with the regeneration of the ATP. An emergency mechanism capable of operating immediately for this purpose is provided by phosphocreatine, which can transfer its high energy phosphate bond to ADP, thus regenerating ATP;

but this mechanism is of limited scope and serves only immediate needs. The main and the ultimate source of the energy for the resynthesis is provided by the process of glycolysis in which glycogen breaks down to pyruvic acid and, in the absence of an adequate oxygen supply, to lactic acid also. This process releases energy for the carbohydrate which is utilized in the building up of new ATP (Fig. 42, 4).

These stages which have so far been briefly described, the breakdown of ATP and its resynthesis, are anaerobic. The final aerobic stage in which pyruvic acid and lactic acid are oxidized by molecular oxygen to yield carbon dioxide and water takes place through the citric acid cycle (tricarboxylic acid cycle) which has already been described in detail on p. 316. This process is accompanied by the synthesis of large amounts of ATP (Fig. 20, 19).

It should be noted that Krebs has recently shown that, *under certain circumstances*, muscular contraction may occur without ATP breakdown.

Carbohydrate Metabolism in Muscle.—Glycogen is synthesized in muscle tissue from the blood glucose in the process of *glycogenesis*. During the process of *glycolysis*, which has already been discussed on p. 311, the muscle glycogen is broken down to pyruvic and lactic acids (Fig. 20, 17).

The importance of lactic acid in muscular contraction was first demonstrated in 1907 by the experiments of Fletcher and Hopkins who found that, if one of a pair of frog sartorii was stimulated in air until it was fatigued, its lactic acid content increased greatly. If the other member of the pair was stimulated in the absence of air, for example in an atmosphere of nitrogen, it became fatigued much more quickly, a larger amount of lactic acid was produced and the glycogen content of the muscle fell considerably. If now oxygen was admitted to the vessel containing such a fatigued muscle some of the lactic acid disappeared and the glycogen content rose.

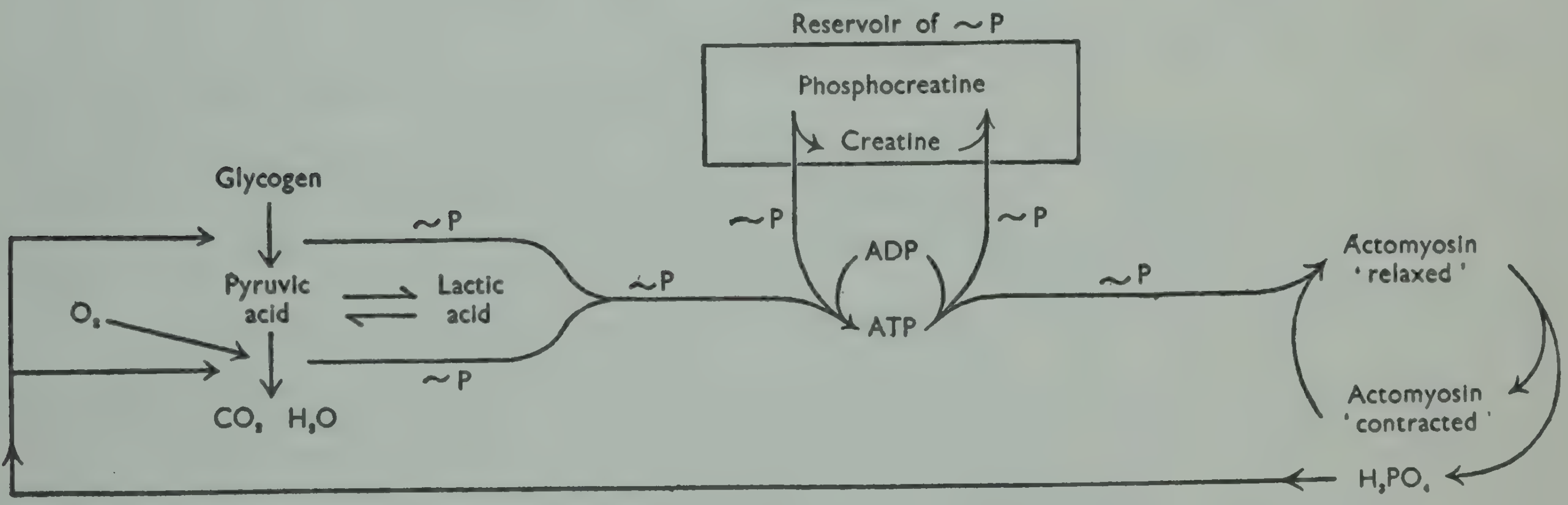


FIG. 42, 4

The chemical basis of muscular contraction. Inorganic phosphate released from ATP by muscular contraction (right) is incorporated by the process of glycolysis (left) into compounds such as ATP and phosphocreatine containing energy-rich phosphate bonds (denoted by ~P). The high bond energy of the ATP is then used in contraction.

(*After* H. N. Munro.)

These simple experiments illustrate some of the fundamental facts of muscle chemistry. Muscular activity results in the utilization of glycogen and the production of lactic acid and its immediate predecessor, pyruvic acid (p. 314). When a muscle which has been fatigued in nitrogen is exposed to air, oxygen is used up, carbon dioxide and water are produced and lactic acid disappears. It used to be claimed that, in frog muscle, four-fifths of the lactic acid produced is reconverted into glycogen and only one-fifth is oxidized to carbon dioxide and water. It is only for the removal of lactic acid (and pyruvic acid) that oxygen is required. The lactic acid produced in muscle is at once neutralized by the buffer systems of the tissues; in mild exercise the oxygen supplied by way of the blood stream may be sufficient for its oxidation in the muscles. In severe exercise, however, lactic acid is produced at a greater rate than it can be dealt with in the muscles and a large proportion escapes from the muscles and reaches the blood, causing a rise in blood lactic acid concentration (p. 799). Here it is neutralized by the buffer systems of the blood and is carried to the liver, where it is converted into liver glycogen. A small amount is often excreted in the urine.

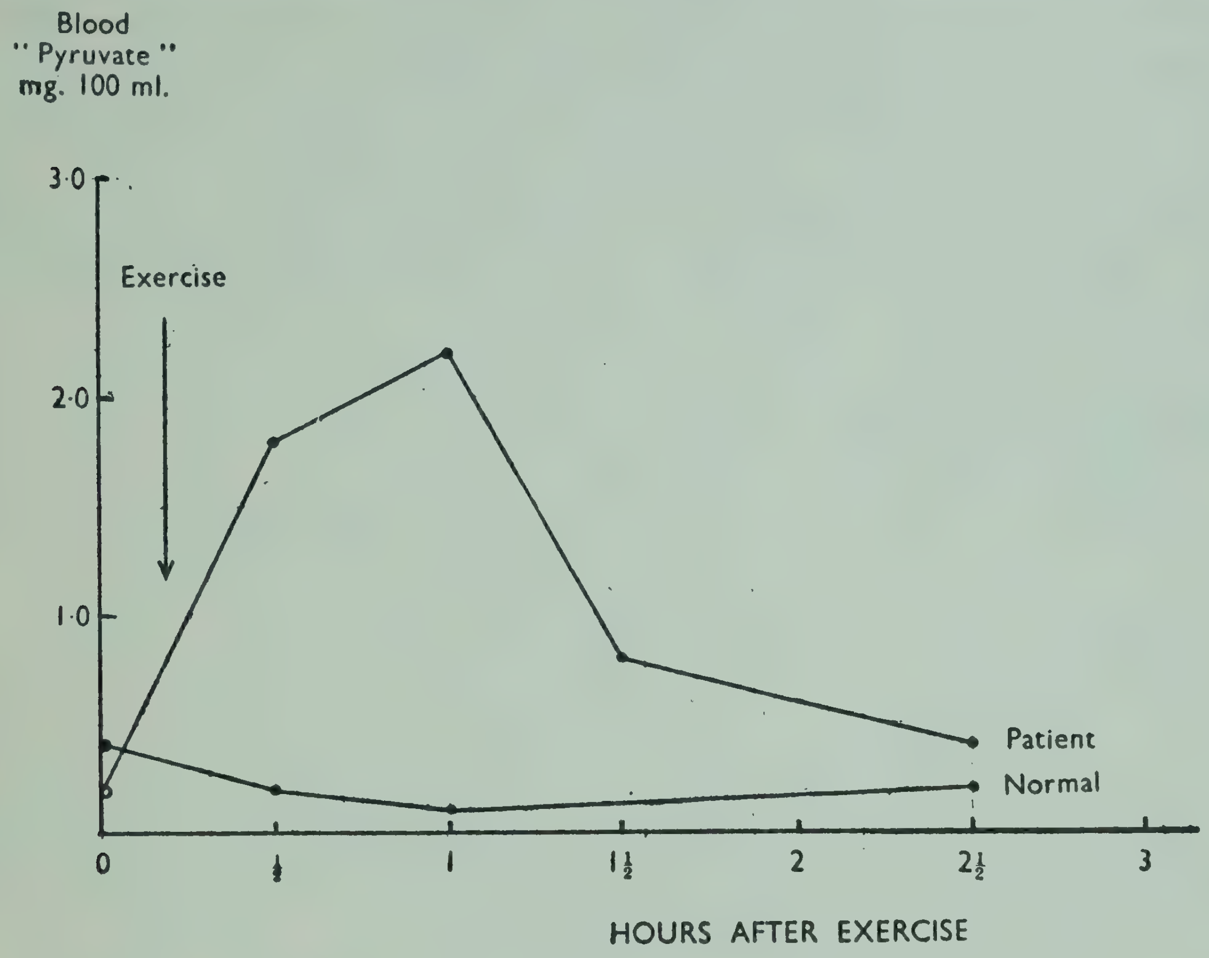


Fig. 42, 5

Blood pyruvate (mg. per 100 ml.) following exercise (7900 ft. Lb. or 1090 kg.m in 5 minutes) in a female aged 58 suffering from vitamin B deficiency, compared with the results in a normal subject.

There are thus essentially three processes involved in muscular contraction:

1. The breakdown of ATP to ADP and phosphoric acid when the myosin fibrils contract, with liberation of energy for contraction.
2. The breakdown of glycogen (glycolysis) to pyruvic and lactic acids by an anaerobic process which takes place quickly to yield energy for the rebuilding of ATP.
3. The aerobic disposal of the pyruvic acid and lactic acid so produced.

The energy liberated by glycolysis and oxidation is used to build up again the ATP which broke down to cause contraction. When the ATP stocks have been replaced and the muscle fibres are again saturated with ATP, glycolysis stops.

Since the process which yields the energy for contraction may be accomplished without oxygen, the body is in fact capable of sudden spurts of activity which require more oxygen than the lungs could take in or the blood could carry during the time of the exercise. The body is then said to be in *oxygen debt* (p. 796). The extra oxygen which is required to oxidize the lactic acid and pyruvic acid may be taken in over a long period after the exercise is over.

Although it is well known that lactic acid may be produced in muscle as the result of repeated contractions, pyruvic acid is to be regarded as a more fundamental product of muscle glycolysis. Normally, of course, it is oxidized at once to carbon dioxide and water, but it may accumulate in the muscle under two sets of circumstances (1) when the rate of glycolysis is very rapid, or (2) if the oxygen available is insufficient for its immediate oxidation. Part of the pyruvate is then converted into lactate which, when the oxygen supply is again adequate, is oxidized back to pyruvate and so to carbon dioxide and water. During the time when the concentration of lactic acid is high in muscle a large proportion may diffuse into the blood by which it is carried to the liver and converted into glycogen.

The oxidation of pyruvate by the citric acid cycle is an elaborate process (see p. 316) which requires among other things the presence of diphosphothiamine (cocarboxylase) derived from thiamine. In the disorder called beri-beri (p. 123), which occurs when there is a severe dietary deficiency of this vitamin, even quite mild forms of exercise lead to an increase in the blood pyruvate because there is not sufficient cocarboxylase to bring about its rapid oxidation. The increase in blood pyruvate following exercise in a case of thiamine deficiency is shown in Fig. 42, 5, compared with the results in a normal individual. This phenomenon has been used as the basis of a test for assessing deficiency of thiamine.

The Aerobic Recovery Phase.—It must not be imagined from what has already been said that muscle is an anaerobic tissue. Muscle tissue is essentially aerobic and, especially during active contraction, shows a high oxygen uptake and carbon dioxide output. Although one stage in the breakdown of muscle glycogen can take place under anaerobic conditions, the energy required for muscular activity is in the long run obtained as the result of oxidation by molecular oxygen carried to the muscle by the hæmoglobin of the blood. The oxygen is transferred to the myohæmoglobin (a form of hæmoglobin occurring in muscle tissue) which has an even greater affinity for oxygen. The oxygen is passed on in turn to the cytochrome (p. 83) of the oxidation enzyme systems in the interior of the muscle cells and is there utilized to bring about the oxidation of pyruvate or lactate or other intermediate products (p. 316).

Rigor mortis.—After death, when the muscles are no longer supplied with oxygenated blood, a gradual stiffening of the muscle tissue occurs which is known as rigor mortis. It was at one time thought that this phenomenon was caused by excessive production of lactic acid but recent work suggests that it is closely associated with a fall in ATP below a certain critical level. Rigor mortis begins much more rapidly if death occurs during severe muscular exertion, or following violent spasms or convulsions, when the ATP stores in the muscle are already low. Its gradual disappearance some 24 to 36 hours after death is probably attributable to softening of the coagulated protein by autolytic ferments.

CHARACTERISTICS OF MUSCULAR CONTRACTION

Among the classical exercises in experimental physiology are experiments on a frog's muscle-nerve preparation. This usually consists of the gastrocnemius muscle with its nerve, the sciatic nerve. When the knee joint is fixed and the muscle tendon attached to a lever writing on a smoked drum, a record of the contraction produced by electrical stimulation of the sciatic nerve or of the muscle itself can be obtained. If a simple pivoted lever is used, as in Fig. 42, 6, the muscle is able to shorten when it is stimulated, the weight on the muscle remaining approximately the same throughout its contraction. This lever is called, therefore, an *isotonic* lever. For some purposes an *isometric* lever is used. A thread from the muscle tendon is fastened to a strong steel spring which has a lever attached to it to magnify its movement. When the muscle is stimulated the tension in it rises but, although the spring is twisted, very little shortening of the muscle is permitted (Fig. 42, 7).

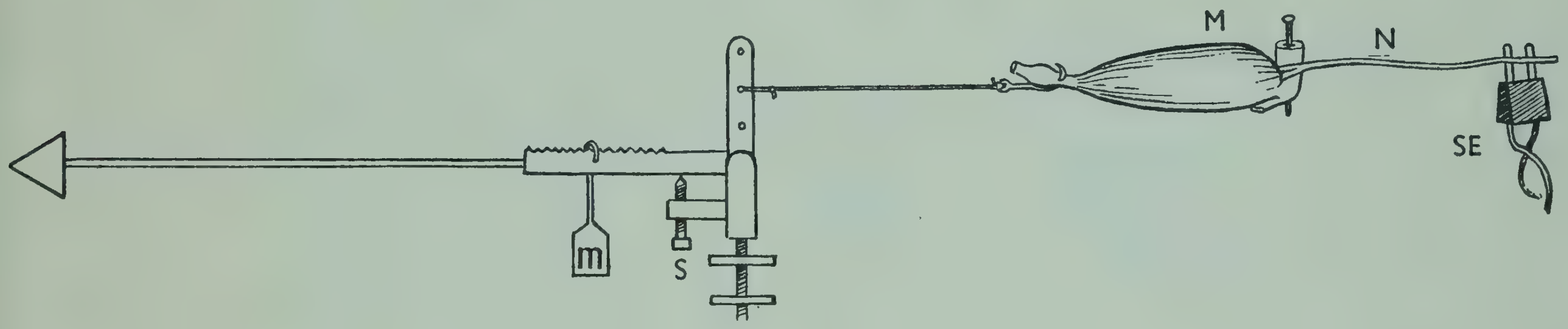


Fig. 42, 6

Isotonic lever. SE, stimulating electrodes. M, gastrocnemius muscle. N, sciatic nerve. S, after loading screw. The weight, m, is hooked over the serrated edge of the lever to adjust the load G. H. Bell (1952), *Experimental Physiology*, 5th edition, p. 23. Glasgow: J. Smith & Son).

When a single electrical impulse is passed directly into a muscle, or into its nerve, the muscle gives a single brief contraction, or *twitch*. The trace obtained when an isotonic lever writes on a rapidly revolving kymograph drum (Fig. 42, 8) shows that the contraction does not begin at the instant at which the impulse is sent in but only after a short *latent period*. The duration of the latent period depends to a large extent on the recording apparatus; with students' apparatus it may be 10 or 20 msec but with optical recording and a lever of low inertia it may be 3 msec, or even less. After the latent period the lever rises for about 50 msec (*contraction period*) and then falls for about 50 msec (*relaxation period*). The twitch recorded by an isometric lever is similar to the isotonic twitch, but both contraction and relaxation are longer in the former. The tension developed by a small frog gastrocnemius muscle may be several hundred grams; the maximum tension is obtained when the excised muscle is attached to the isometric lever in such a way that the length of the muscle fibres is approximately the same as their resting length in the body (Fig. 42, 9).

The tension developed in an isometric twitch depends on the strength of the electrical stimulus applied to it. The weakest stimulus which has an effect is said to be a *threshold* or *liminal* stimulus. As the stimulus is increased above this value, provided that the muscle is not fatigued, the tension produced in a twitch is greater, up to a maximum value which is not

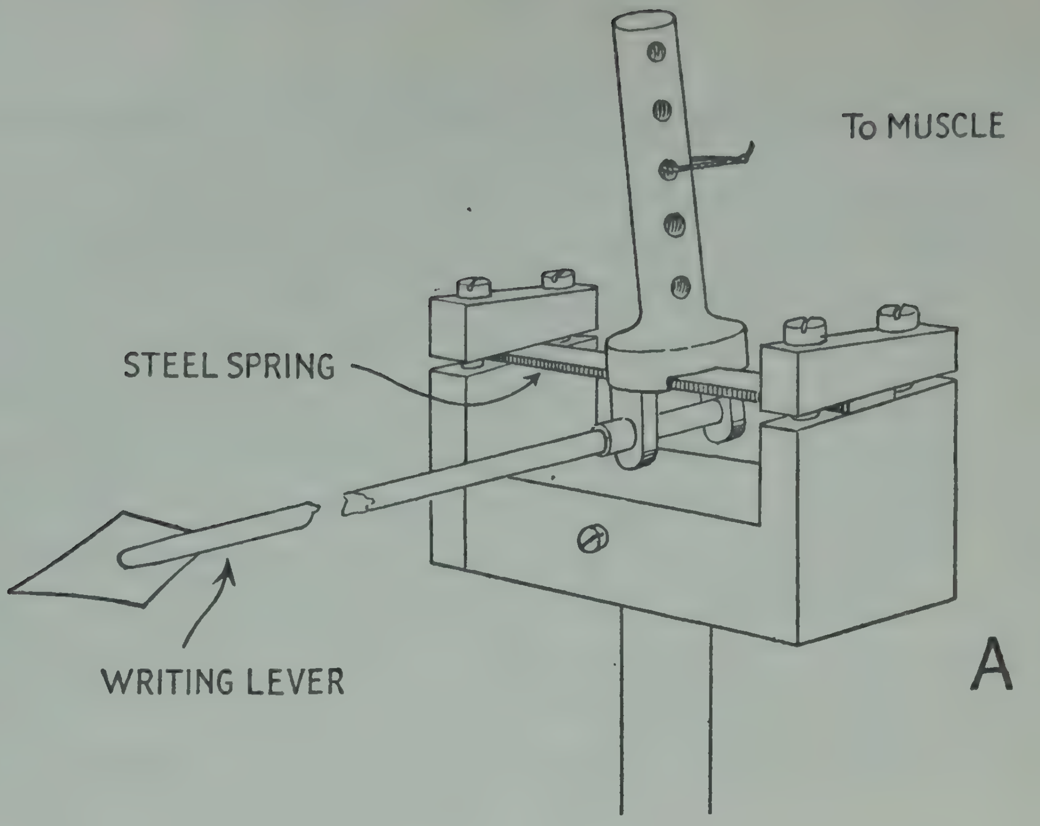

To MUSCLE
STEEL SPRING
WRITING LEVER
A
ISOMETRIC
LEVER
M
C
N
B

FIG. 42, 7

Isometric lever. A shows the detail of construction and B the general arrangement of the apparatus when a frog muscle is used.

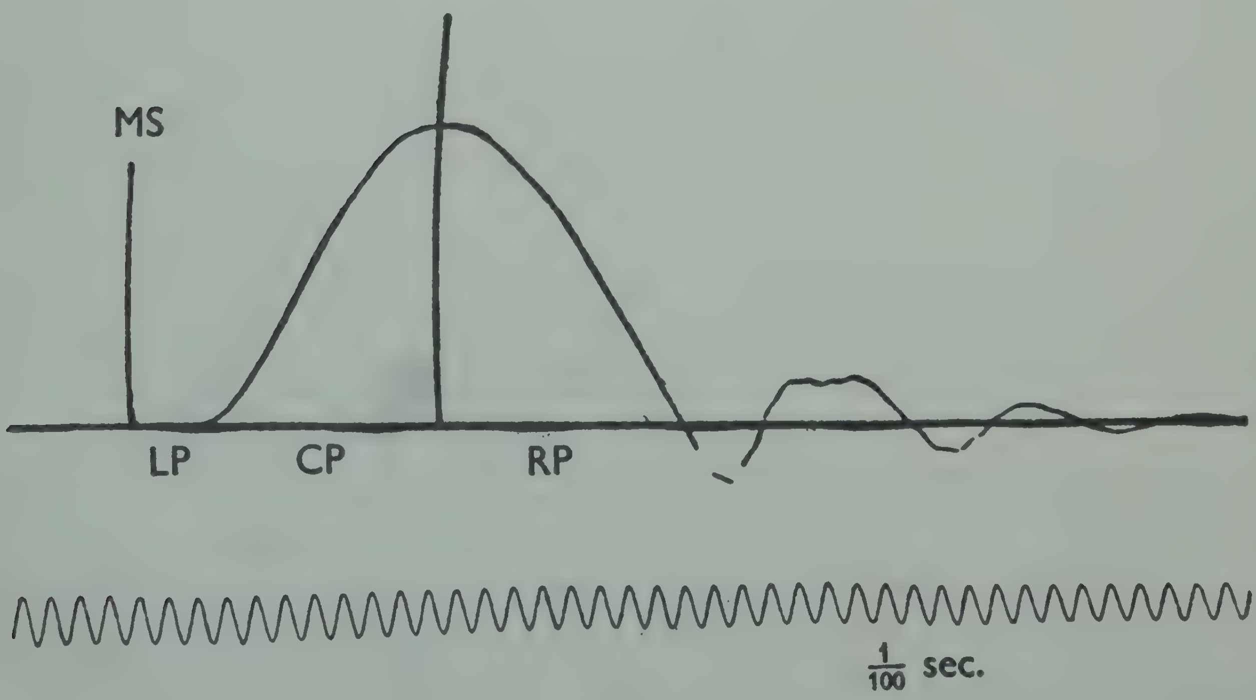


FIG. 42, 8

Record of isotonic twitch. LP, latent period. CP, contraction period. RP, relaxation period. MS, moment of stimulation. The waves at the end of the trace are due to the lever bouncing on the after loading screw (see Fig. 42, 6) (G. H. Bell (1952), *Experimental Physiology*, 5th ed. p. 25. Glasgow: J. Smith & Son).

exceeded if the stimulus is made *supramaximal.* The gradation in the strength of the twitch is not due to a graded response by the individual fibres of the muscle but to variation in the number of muscle fibres in action. By examining muscle fibres under the microscope when they are stimulated with a very small electrode it is found that each fibre contracts to its maximum or not at all, that is, it obeys the 'all-or-nothing' law (Fig. 42, 11). Each individual muscle cell, if it responds at all to a stimulus, gives a maximum contraction, and this maximum depends on the particular conditions of the experiment but not on the strength of the stimulus. The extent of the response may be influenced by the initial length of the cell or by fatigue but these factors do not affect the all-or-nothing character of the response. This generalization applies to all excitable tissues. The strength of the contraction of a skeletal muscle depends on the number of its fibres which are in action. A threshold stimulus excites only a few cells; stronger electrical stimuli spread further into the muscle and cause more cells to contract. This behaviour is quite different from that of cardiac muscle in which, since all the cells are in functional continuity (p. 425), the response cannot be adjusted by altering the number of cells in action.

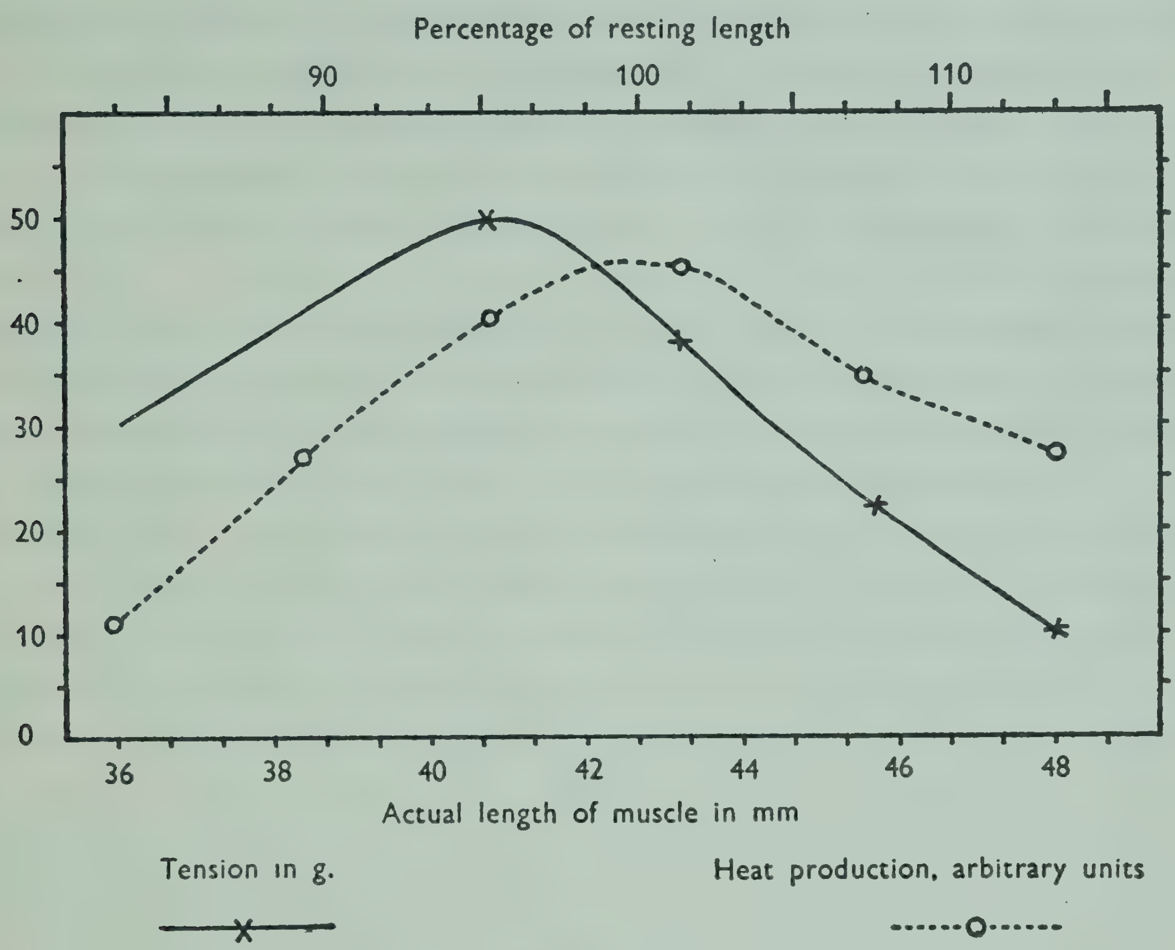


FIG. 42, 9

Graph to show the heat production (○) and the tension (×) developed by two frog sartorius muscles on stimulation by single maximal induction shocks. Before each stimulation the position of the isometric lever was adjusted so that the muscles were either at their natural length or at a length a little above or below this value. Maximal heat production occurred when the muscles were at their natural length, while maximal tension was reached when the muscles were slightly shorter. The maximal value of the ratio Tension/Heat production is found when the muscles are at their natural length; at high tensions the ratio diminishes rapidly (C. L. Evans & A. V. Hill (1914), *J. Physiol.* **49,** 12).

If two electrical impulses with only a short interval between them are applied to a muscle it responds by two contractions provided that the stimuli are more than 5 msec apart. When the second impulse is sent in during the contraction phase produced by the first impulse, a second contraction is superimposed on the first so that the total height of the contraction is much greater than that elicited by a single impulse. The shape of the curve recorded depends on the time interval between the two stimuli; this effect is illustrated in Fig. 42, 12.

When the time between the two stimuli is shorter than 5 msec in the frog's muscle, or about 2 msec in mammalian muscle, the second stimulus falls in the *absolutely refractory period* following the first stimulus and is without effect, that is the contraction produced is the same as that obtained with the first stimulus alone. The refractory period, which is about the same length as the latent period and exceedingly short when compared with the contraction period, is occupied by the process of repolarization which must be complete before the muscle can respond fully to a second stimulus (see p. 739).

The shortness of the absolutely refractory period of skeletal muscle accounts for the summation of contractions and it permits a sustained contraction or *tetanus* to occur when a rapid series of stimuli is applied to the muscle. The tension developed in a muscle in tetanus may be several times that developed at the height of a single twitch. The character of the record obtained (Fig. 42, 13) depends on the interval between successive stimuli. The rate needed to produce a smooth curve or *complete tetanus* (D in Fig. 42, 13) varies from muscle to muscle, and is obviously lower in a slowly acting muscle. Low rates of stimulation produce an *incomplete tetanus* or *clonus* (Fig. 42, 13, A, B, and C).

If the frog's muscle is warmed, say to 25° C, all the metabolic processes are hastened and the viscosity of the muscle cytoplasm is reduced, with the result that the latent period is shortened and the lever rises and falls more smartly. If the muscle is cooled to, say, 5° C, all the chemical activities are slowed down and the twitch is prolonged by the increase in viscosity. As one might expect from the higher body temperature a twitch produced by a mammalian muscle is rather shorter in duration than that produced by a frog muscle, the exact time depending on the actual muscle (see upper curve of Fig. 42, 19).

MOTOR UNIT.—Sherrington introduced the term *motor unit* in 1925 in these words: "The muscle with its nerve may be thought of as an additive assembly of motor units, meaning by motor unit an individual motor nerve fibre together with the bunch of muscle fibres it innervates." It seemed at that time simple enough in principle to find the number of muscle fibres in a motor unit. After section of the dorsal roots of the spinal cord the sensory fibres degenerate but the fibres emerging from the ventral roots remain intact. A count of the fibres remaining in the trunk of the motor nerve was made as far away from the muscle as possible because the nerve fibres branch repeatedly especially after the nerve has entered the belly of the muscle. When the number of cells in the muscle supplied was counted it was found that in the cat the innervation ratio of limb muscles was of the order of 1 in 150; this appeared to show that one motor nerve cell of the anterior horn of the spinal cord supplied a bunch of 150 muscle cells. Recent discoveries, however, show that the problem is not capable of such a simple solution. A considerable fraction—varying from muscle to muscle but of the order of one-third—of the efferent fibres leaving the ventral roots supply muscle-spindles and are not concerned with motor units. On this basis the number of the muscle cells in a motor unit must be increased by about 50 per cent. over the original estimate. It is likely, however, that in the case of muscles used for skilled movements such as those of the hand the number is relatively small. The tension developed by a single motor unit of the flexor digitorum longus of the cat has been measured by stimulating a single motor nerve fibre when the muscle was connected to an isometric lever. Values as high as 42·5 g. per motor unit were obtained. A further complication is that the original assumption that no muscle cells receive nerve fibres from more than one nerve cell is incorrect. In some muscles it has been shown that if the motor nerve is dissected into two roots simultaneous stimulation of the two roots gives little more tension than stimulation of a single root. This indicates a considerable overlap in the innervation. Furthermore, stimulation of nerves arising from quite separate points of the central nervous system may activate a single muscle cell as shown by a microelectrode inserted into it. In view of these recent ideas on polyneuronal innervation the original conception of the motor unit requires some modification. The use of the word 'bunch' in Sherrington's description of a motor unit suggests that all the muscle cells in the unit are close together; this is not necessarily so. Several investigators have found that a single nerve fibre may send branches to widely separated cells within a single muscle.

Voluntary Muscular Contraction.—When the arm is flexed against resistance the biceps muscle appears to increase in size but this is merely an alteration in its shape and the change in volume is negligible. There is, in fact, a very slight shrinkage on contraction. When a muscle contracts alterations of the polarized membranes of the muscle cells occur, just as in nerve when it is conducting (p. 737), and these can be detected by placing electrodes on or in the muscle and leading them to an oscillograph. If simultaneous records are made of the electrical variations in a muscle, and of the tension exerted by it when it is stimulated through its nerve, it is found that the electrical variations occur a short time before the mechanical.

A voluntary contraction of a muscle may be a weak or a powerful sustained contraction. The experiments already described suggest that grading could be brought about either by variation in the number of motor units in action at any one instant, or by variations in the rate of discharge of the motor cells in the spinal cord. If two small electrodes placed very close together are applied to a muscle the potentials due to a few muscle cells can be recorded and it is found that both methods of grading occur simultaneously. It can be seen from Fig. 42, 14, that as the voluntary contraction becomes more powerful the muscle cell responsible for the main spike on the record discharges more frequently, but during the most powerful contraction (D) smaller spikes due to the activity of adjacent cells also appear. The irregularities shown in the electrical record of a few motor units are not of course seen in the mechanical record of the total contraction of the whole muscle. A smooth voluntary contraction is the algebraic sum of the contractions and relaxations of a large number of motor units, each of which is contracting irregularly, asynchronously and intermittently. Since each muscle cell obtains in this way a period of rest a voluntary contraction shows much less fatigue than an experimental, electrically produced tetanus in which all the cells involved are in sustained contraction.

Muscle tone, the continuous slight tension in which all the skeletal muscles are maintained in health, is a reflex phenomenon the nature of which is considered on p. 841. When the spinal cord or the afferent or the efferent nerves of a muscle are destroyed, tone is lost and the muscle becomes flaccid. Tone consists presumably of a minimal sustained muscular contraction produced, at any one time, by a few motor units. If the groups of motor units responsible for tone change periodically the maintenance of tone without fatigue can be accounted for. This conception of tone is however difficult to reconcile with the finding that some normal skeletal muscles at rest do not show any electrical activity (Fig. 42, 14, A).

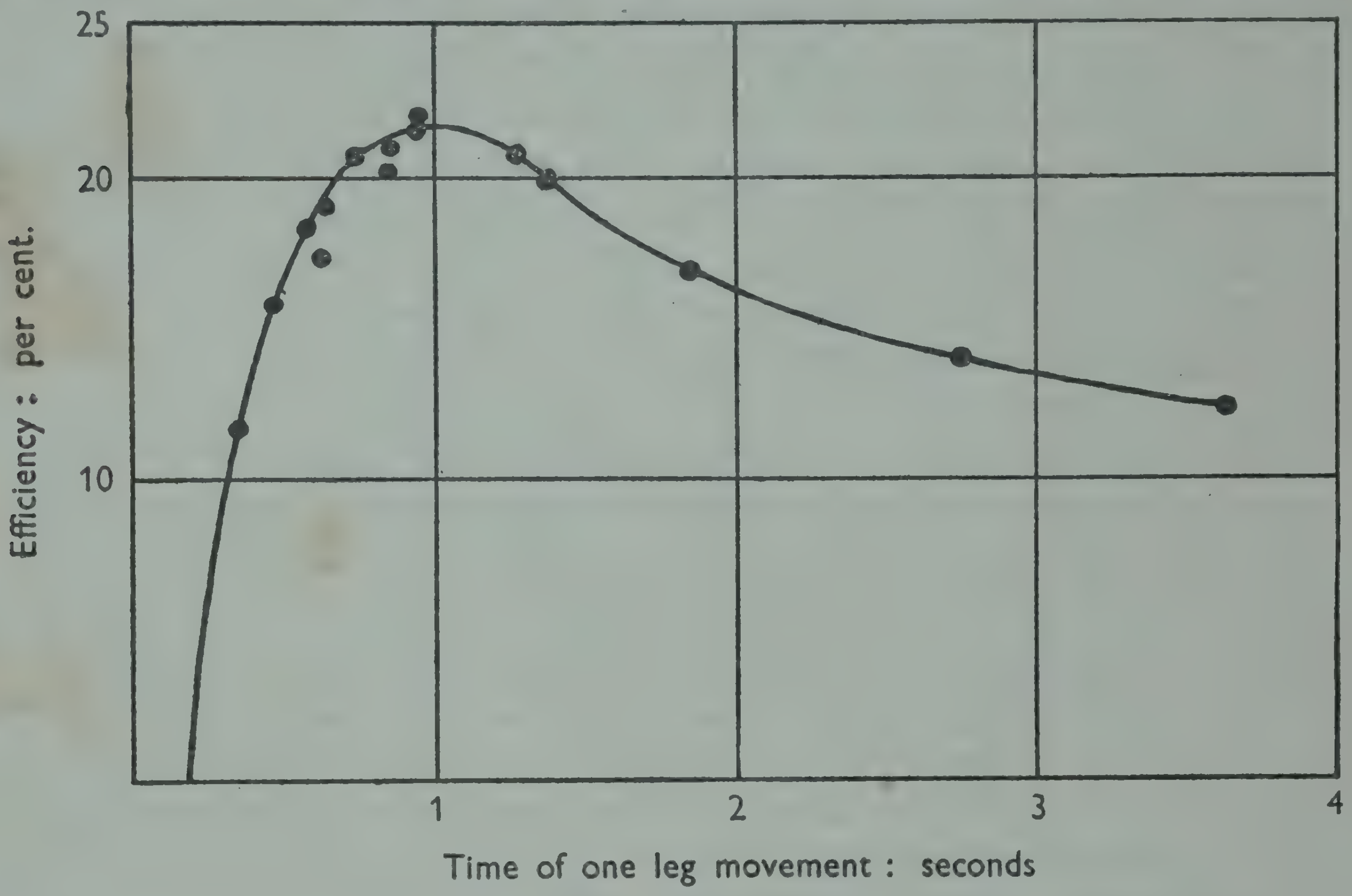


Fig. 42, 10

The diagram shows the net efficiency of a subject pedalling a bicycle ergometer. The graph gives the relationship between the efficiency and the time of one leg movement. The curve is drawn through the observed points. The optimum time of one foot movement (half a pedal revolution) is 0·9 of a second (S. Dickinson (1929). The efficiency of bicycle-pedalling. *J. Physiol.* **67**, 246).

Optimum Rate of Working.—It is within ordinary experience that there is an optimum rate for carrying out any task requiring muscular effort. If the rate is too high fatigue sets in and the work has to be stopped; if the rate is too low energy is wasted. Hill measured the metabolic cost of climbing steps and found that the optimum rate was one step in one to one and a half seconds, that is to say, approximately one muscular movement per second. Dickinson came to a similar conclusion in her experiments on bicycle-pedalling (Fig. 42, 10). Although the optimum rate found in any one person may be a reflexion of the capabilities of his circulatory and respiratory systems it is dependent to some extent on the properties of the muscles themselves.

A muscle can be thought of as behaving like a thin-walled rubber tube containing a viscous fluid. In a very rapid contraction all the work is done against the viscous resistance of the muscle and none is available for useful external work. The maximum theoretical work which can be obtained

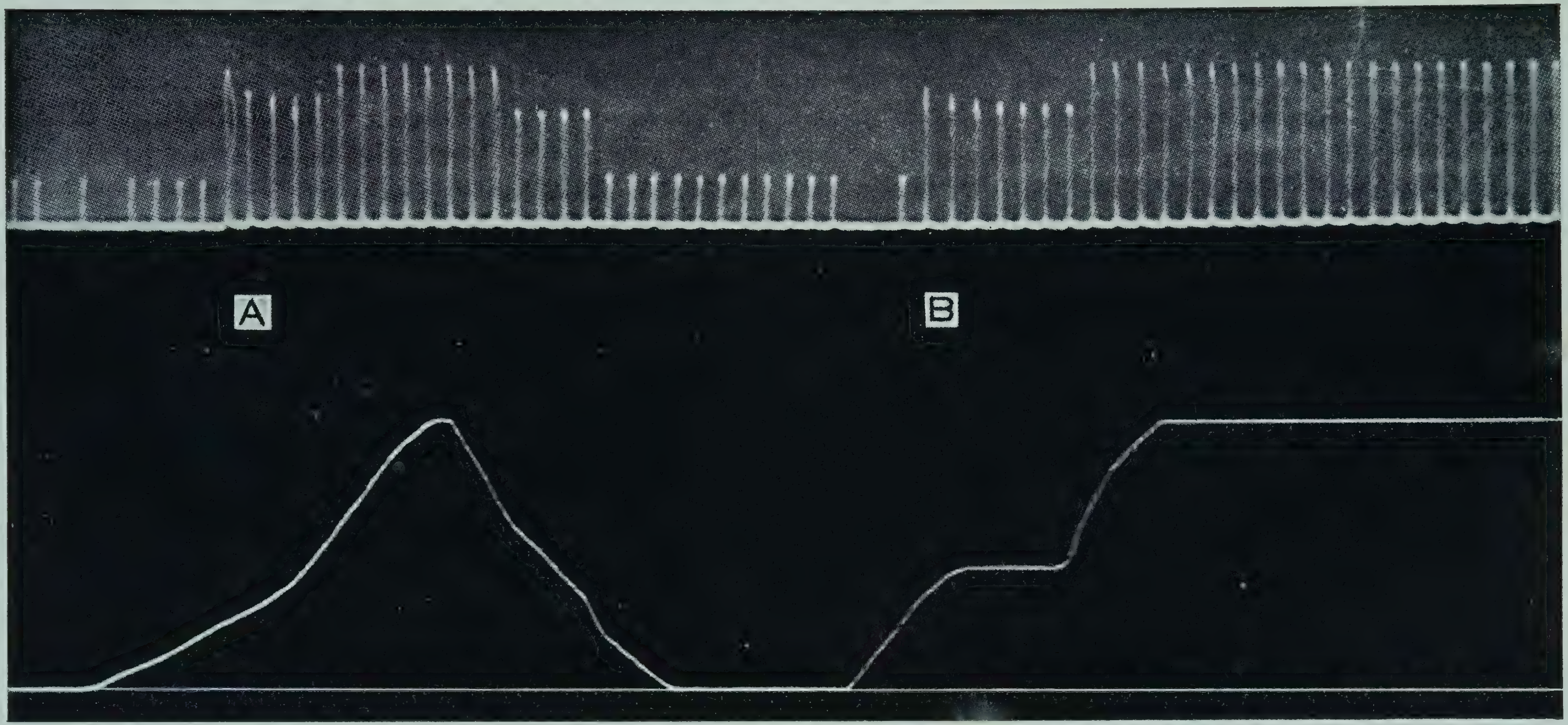


FIG. 42, 11

Quantal phenomena in sartorius of Rana Pipiens. The lower line is the record of the movement of the secondary coil of an induction coil; a rise of the line indicates that the primary and secondary coils were approaching one another. The upper trace gives the movement, photographed through a microscope, of a droplet of mercury lying on the sartorius muscle of a frog which was stimulated through a pore electrode 7 μ in diameter. The muscle was stimulated once per second. The rise in strength of stimulus at A caused a quantal increase (not a graded response) in the muscular contraction. Just before B several responses are missing owing to fatigue. From B onwards there is a continuous decrement due to fatigue (F. H. Pratt & J. P. Eisenberger (1919). *Amer. J. Physiol.* **49**, 45).

[*Facing page* 784

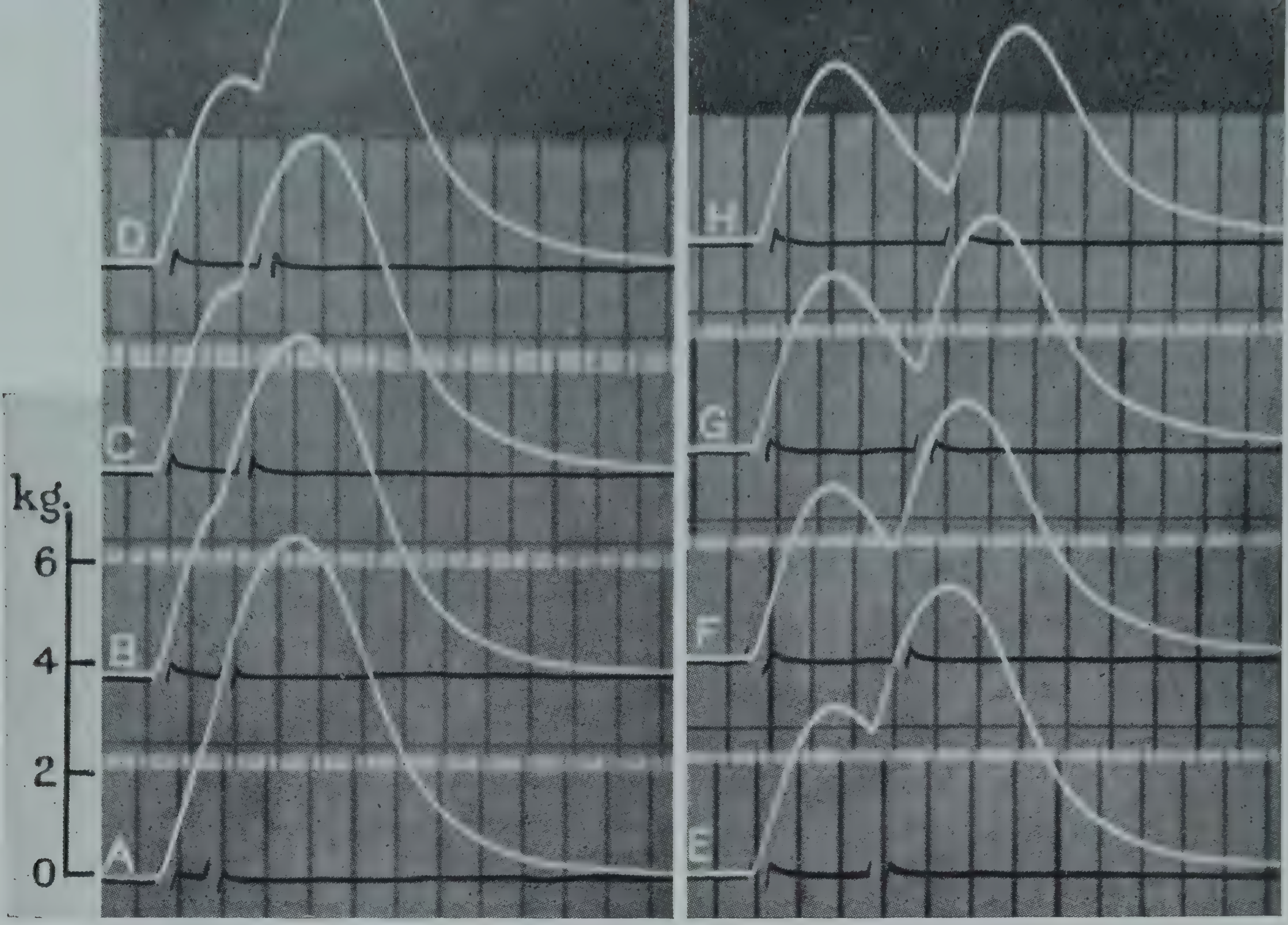


FIG. 42, 12

Mechanical response of median head of gastrocnemius of a cat to double shocks to peroneal nerve. The mechanical response (white line) was recorded by an isometric frictionless mirror myograph. The electrical response (dark line) was recorded by a string galvanometer. A Lucas pendulum was used to give two maximal break shocks in quick succession to the nerve. Temperature, 34° C. Interval between stimuli: A, 24 msec; B, 32 msec; C, 40 msec; D, 48 msec; E, 57 msec; F, 69 msec; G, 77 msec; H, 88 msec. Time is shown by vertical lines 20 msec apart (Sybil Cooper & J. C. Eccles (1930), *J. Physiol.* **69**, 379).

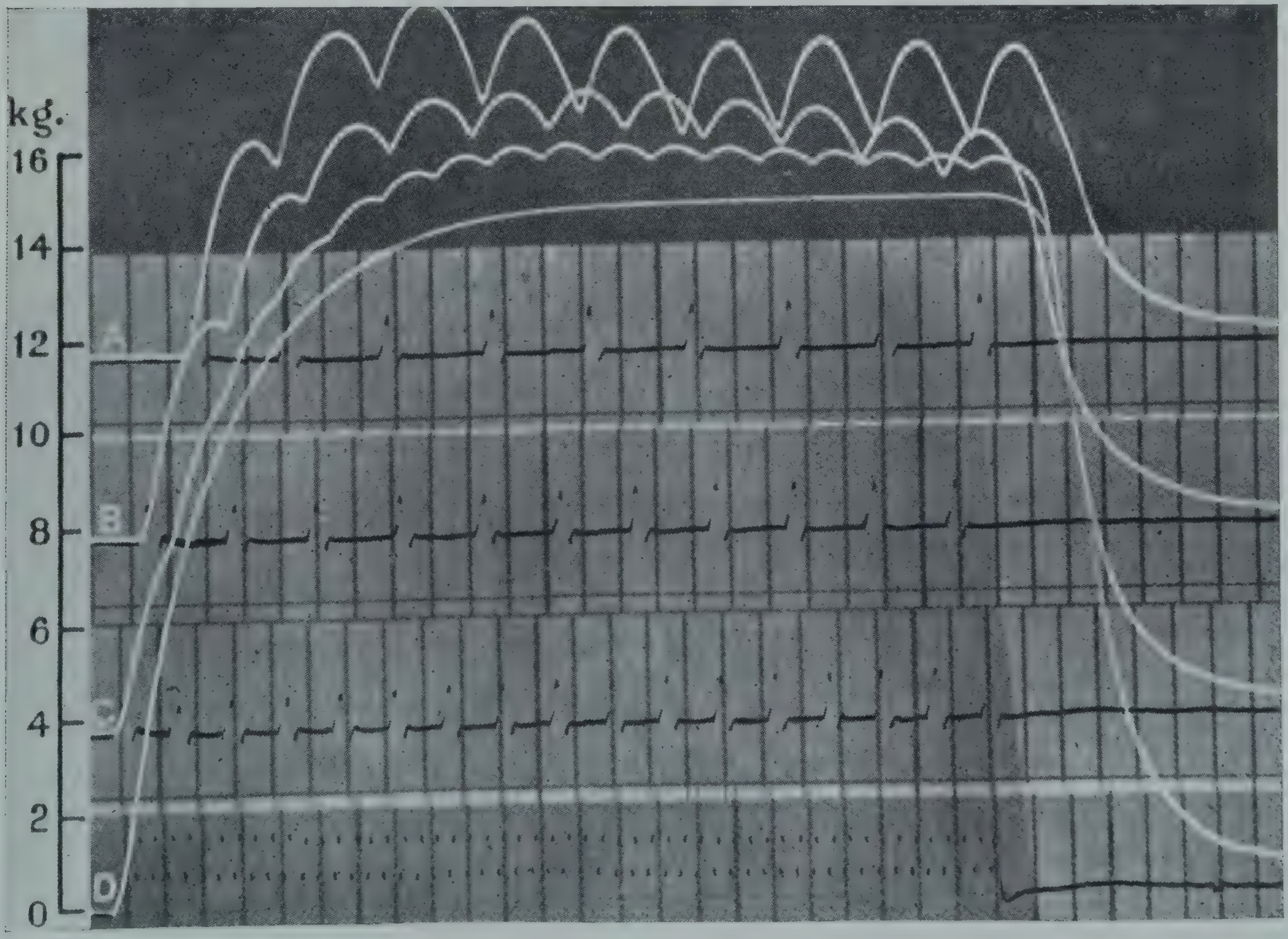


FIG. 42, 13

Mechanical and electrical responses of median head of gastrocnemius of a cat to tetanic stimulation of popliteal nerve by a neon tube stimulating device. Temperature, 33° C. Rate of stimulation: record A, 19 shocks a second; record B, 23·5 shocks a second; record C, 35 shocks a second; record D, 315 shocks a second. Time is indicated by vertical lines 20 msec apart (Sybil Cooper & J. C. Eccles (1930), Isometric responses of mammalian muscles. *J. Physiol.* **69**, 383).

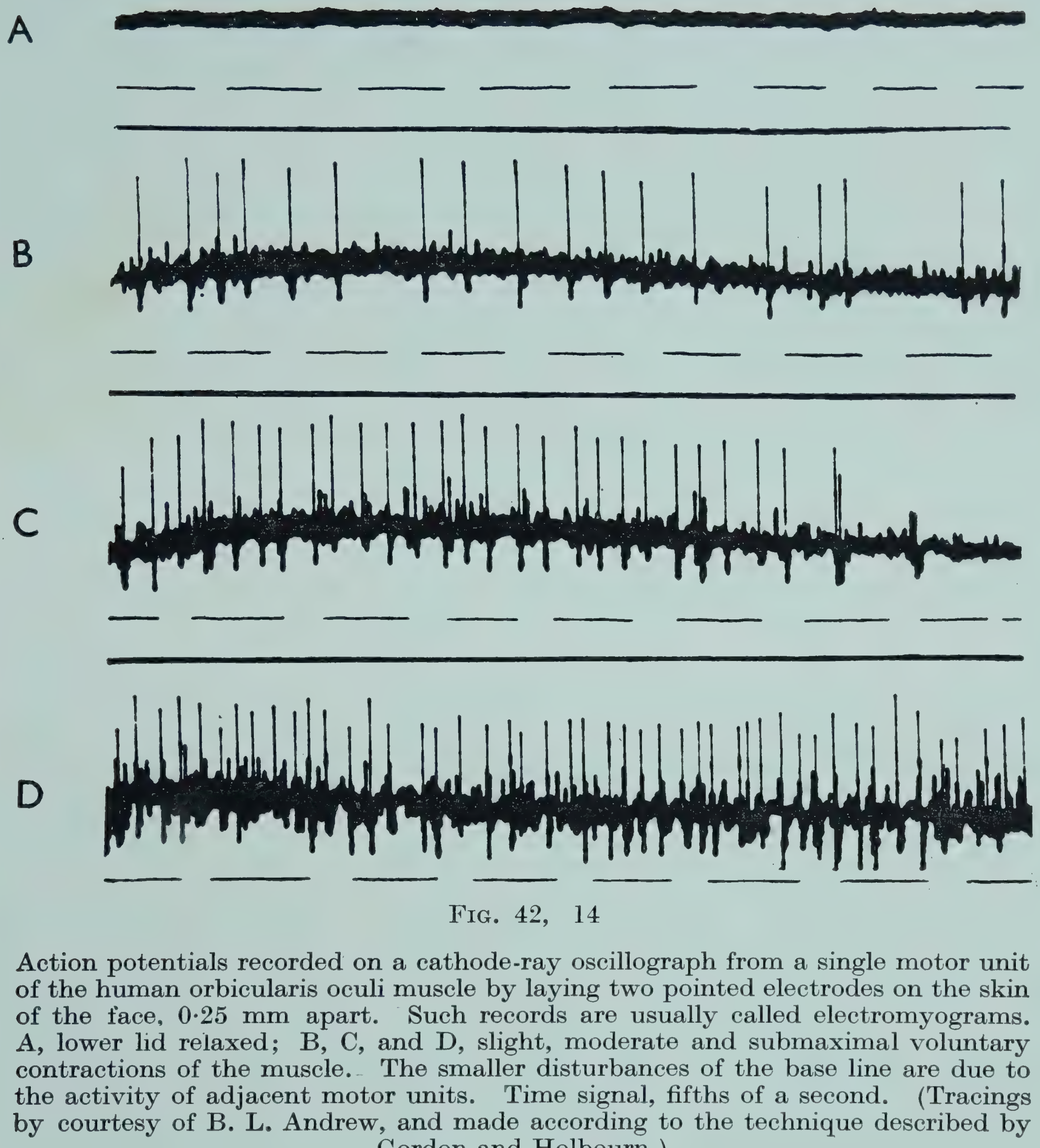


Fig. 42, 14

Action potentials recorded on a cathode-ray oscillograph from a single motor unit of the human orbicularis oculi muscle by laying two pointed electrodes on the skin of the face, 0·25 mm apart. Such records are usually called electromyograms. A, lower lid relaxed; B, C, and D, slight, moderate and submaximal voluntary contractions of the muscle. The smaller disturbances of the base line are due to the activity of adjacent motor units. Time signal, fifths of a second. (Tracings by courtesy of B. L. Andrew, and made according to the technique described by Gordon and Holbourn.)

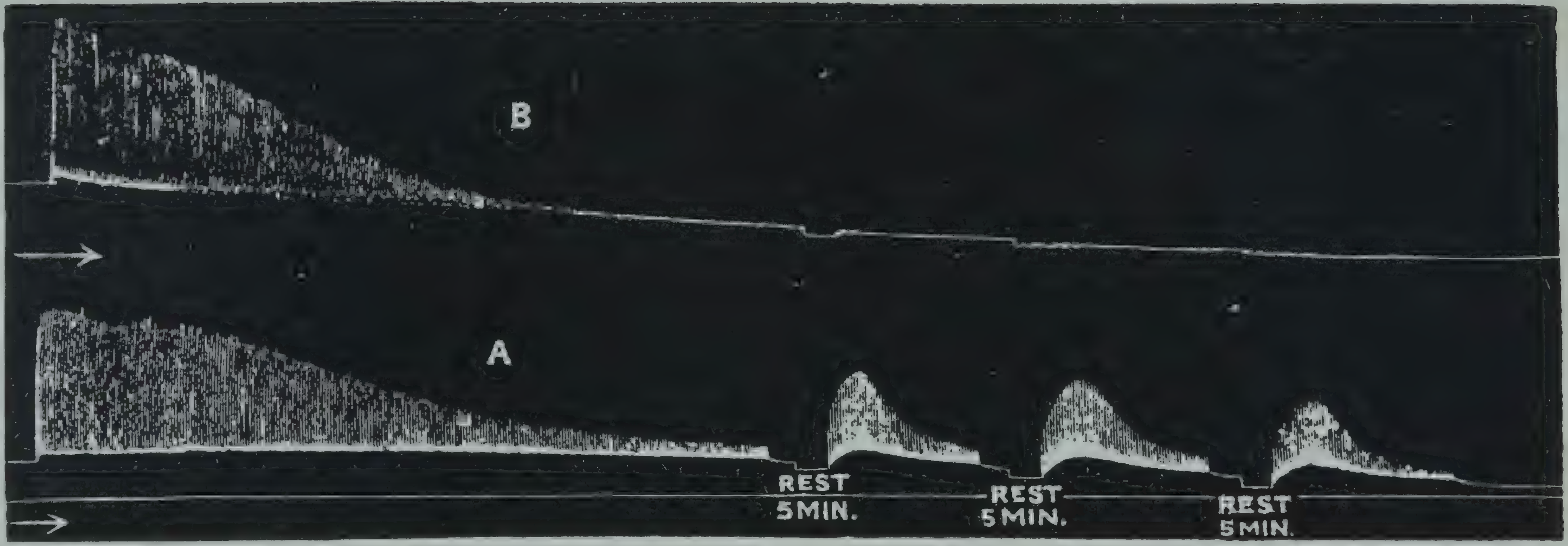


Fig. 42, 15

The two frog sartorius muscles **A** and B attached to an isotonic lever were stimulated with maximal induction shocks through electrodes on the two ends of the muscle once every second except during the rest periods, when the kymograph drum was stopped. Muscle A was exposed to oxygen, while muscle B was kept in nitrogen (W. M. Fletcher (1902), Oxygen and muscle metabolism, *J. Physiol.* **28**, 492).

from a muscle is equal to the external work actually obtained plus the work done by the muscle against the viscous resistance to change of form. The importance of viscosity has been investigated quantitatively by A. V. Hill as follows. A heavy flywheel mounted in ball bearings is used to provide the inertia against which the biceps muscle has to work (Fig. 42, 16) by pulling on a flexible wire wound round a pulley. When the arm is flexed the flywheel is accelerated up to a maximum speed, the wire is unwound and slips off the pulley and the wheel continues to revolve after the contraction is over. The work done is proportional to the square of the maximum speed in revolutions per minute. By using different pulleys different equivalent masses can be obtained, for the smaller the pulley used the greater the equivalent mass. When a maximum muscular effort is made with each size of pulley the curve of work done against the time occupied in the flexion of the arm is as shown in Fig. 42, 17. The maximum external work is obtained when the flexion occupies at least two seconds, that is, when the contraction is very slow. In a quick contraction lasting 0·25 seconds no external work is done at all. This experiment is somewhat artificial because it deals with only one movement which is not repeated, and it must not be concluded that the greatest efficiency is obtained by working infinitely slowly. If muscular movements are carried out very slowly efficiency falls off, because more prolonged actions require additional energy for maintenance of posture, for instance even to lift a limb and keep it lifted requires energy, but no external work is done while the position of the limb is being maintained. The optimum duration of a skeletal muscular movement is about 1 sec and this corresponds to the rate of bicycle-pedalling (70 per min) which has been found by several workers to give the highest efficiency in their subjects (Fig. 42, 10).

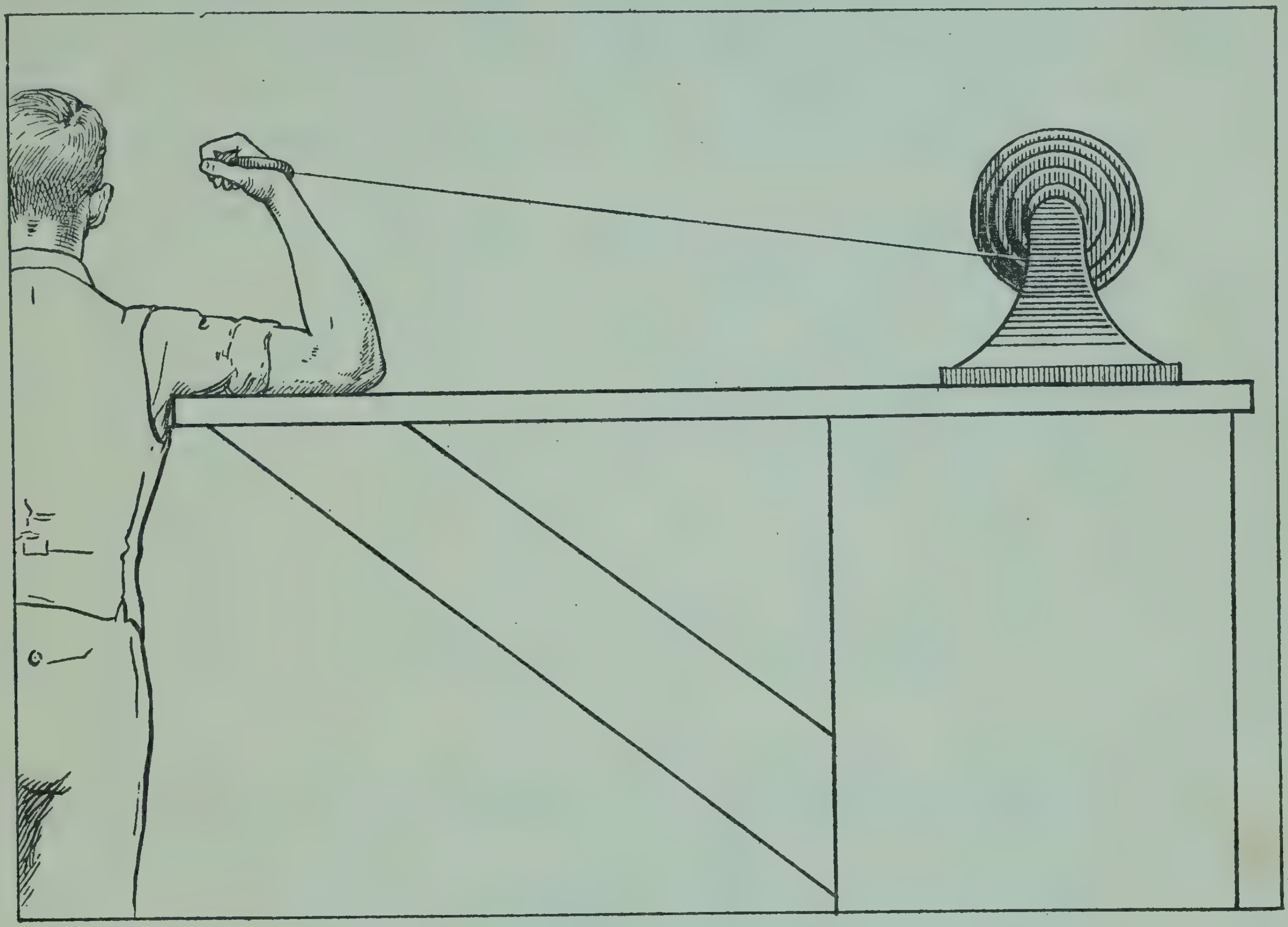


FIG. 42, 16

Inertia flywheel for measuring the work done by human muscles (A. V. Hill (1922), *J. Physiol.* **56,** 22).

Viscosity must not be thought of solely as a hindrance to muscular movement, for the viscosity of muscles acts as an automatic brake preventing too rapid movement. A. V. Hill points out that if man's muscle were less viscous he would, if he indulged in vigorous exercise, be very apt to break bones and tear tendons. If a muscle is allowed to contract the maximum tension it can develop is much less than the isometric tension. For this reason also harmful increases of tension are less likely to occur in the living body where isometric contractions rarely occur.

FATIGUE.—When a muscular effort is repeated frequently the subject begins to tire. This fatigue is due partly to the properties of muscle tissue itself and partly to the properties of the central nervous system. When a frog's muscle is repeatedly stimulated through electrodes placed directly on it the contractions after a time become smaller, until finally the muscle is unable to respond (Fig. 42, 15). Before this stage is reached, however, the muscle relaxes more and more slowly after each twitch; if the time between successive twitches is short, say half a second, the fatigued muscle is not able to relax to its original resting length as shown in trace A of Fig. 42, 15. When stimulation ceases the muscle gradually regains its excitability. Oxygen is essential for recovery, for a muscle, stimulated to complete fatigue and left in an atmosphere of nitrogen, shows no recovery (Fig. 42, 15, B). Fatigue in such a muscle-nerve preparation may be supposed to be due to the accumulation in the muscle of metabolites, perhaps lactic acid, which in some way clog the machine. The oxidative removal of these substances allows the machine to work once more.

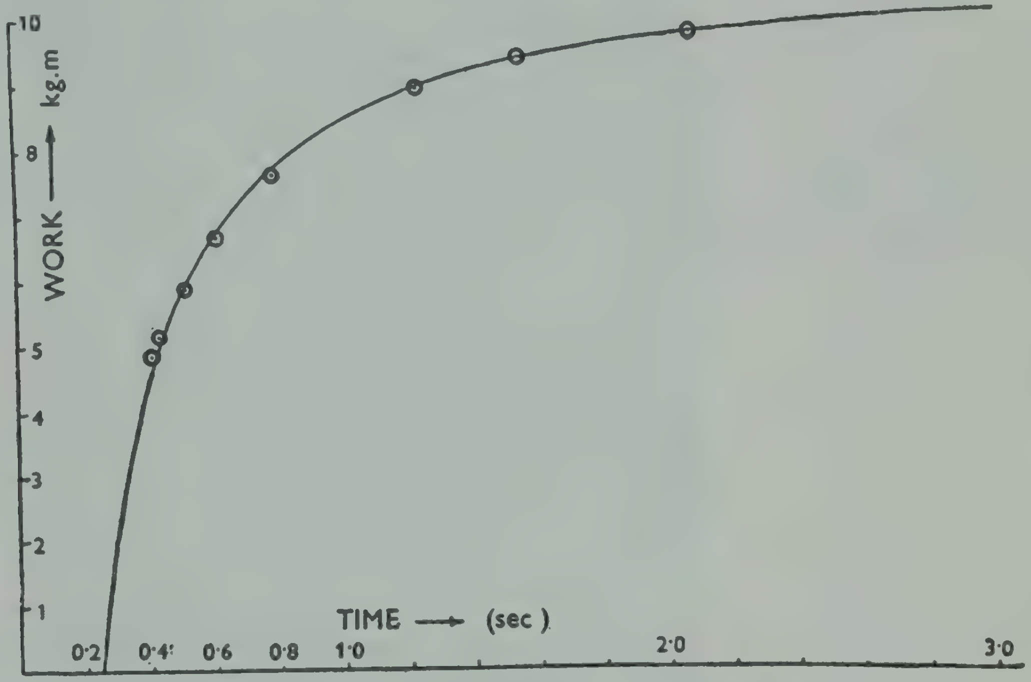


FIG. 42, 17

Relation between maximum work and duration of contraction (A. V. Hill (1922). Work of muscles. *J. Physiol.* **56,** Fig. 4, p. 30).

A muscle of a decerebrate cat yields approximately the same tension whether stimulated directly or through its nerve but, when a rapid series of stimuli is used to produce a tetanus, the tension usually, though not invariably, declines slightly more rapidly when the preparation is stimulated through the nerve. A change over from nerve stimulation to direct stimulation produces a rise in tension, as shown in the upper curve of Fig. 42, 18. Muscle cells which respond to direct stimulation fail to respond to the nerve volleys and hence we conclude that there is a neuromuscular block.

The refractory period of some peripheral part of the system, probably the neuromuscular junction, has increased, so that the muscle cells can respond only to a certain number of the impulses which reach them. After ten seconds of stimulation of the nerve at the rate of 100 per second in the experiment illustrated in Fig. 42, 18, the tension fell to 50 per cent. of the initial value but, since direct stimulation of the muscle at 25 per second produced the same tension, only one-quarter of the cells of the muscle were responding at any one moment. Each individual muscle cell when it came into action must therefore have been contributing very much more to the tension than it did initially. Accordingly the development of neuromuscular block is not responsible for the fall of tension which occurs as the muscle-nerve system fatigues.

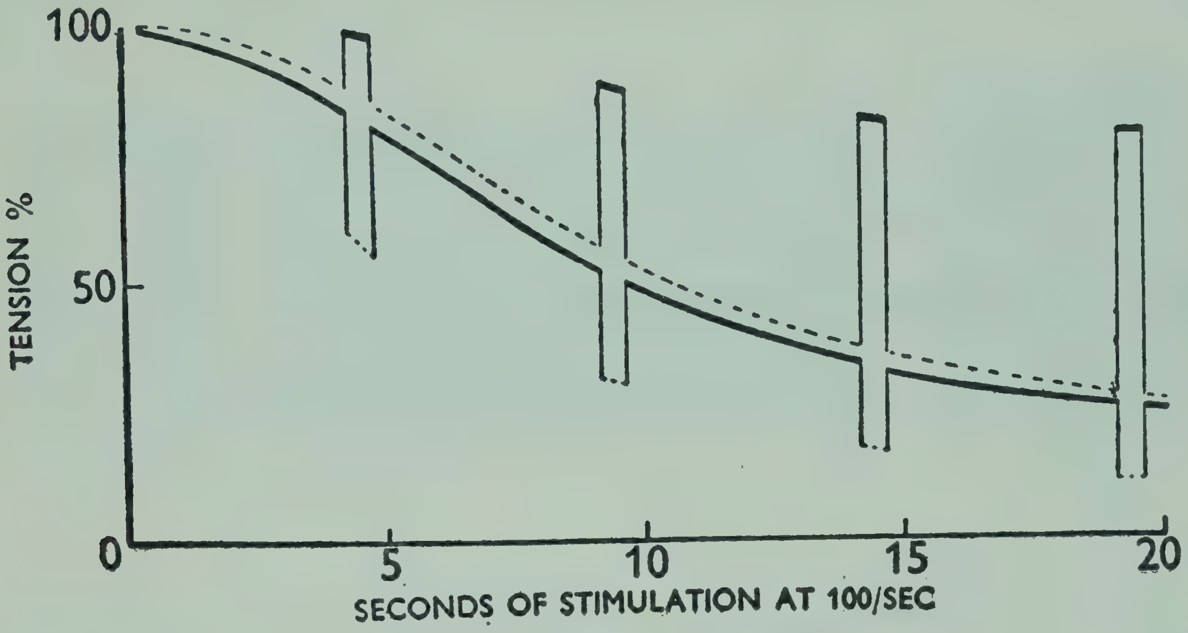


FIG. 42, 18

Changes in tension of the tibialis anterior of a decerebrate cat when nerve stimulation is exchanged for direct stimulation at the same frequency. Upper curve, nerve stimulation interrupted by direct stimulation. Lower curve, direct stimulation of the same muscle interrupted by nerve stimulation.

—— Tension due to direct stimulation.

- - - - Tension due to nerve stimulation.

(G. L. Brown & B. D. Burns (1949), *Proc. Roy. Soc.* B, **136,** 186, Fig. 2.)

By passing a rapid series of electrical stimuli through the foot of a spinal frog a reflex withdrawal of the foot is obtained. When the stimulation is continued the reflex is gradually fatigued and the foot drops back slowly to its original position. If at this stage the nerve of the flexors is stimulated a contraction again occurs. The cessation of the reflex contraction cannot, therefore, be due to failure of excitability of the muscle or of the nerves, but it must be due to changes in the spinal cord. A similar phenomenon can also be shown in man by making a subject contract his forearm muscles repeatedly to lift a weight. When he is no longer able to lift the weight voluntarily, electrical stimulation of the motor nerve through the skin produces a powerful contraction.

The fatigue experienced in severe exercise is due to events at the synapses in the central nervous system. This may be regarded as a protective mechanism, since central fatigue is manifest before there is any block at the neuromuscular junction and long before the muscle itself is incapable of contraction. It should be kept in mind that the word fatigue has also a psychological meaning; fatigue may, for example, arise through lack of interest in a routine and boring task. This is very different from the physiological meaning of fatigue.

MYONEURAL JUNCTION.—If a skeletal muscle is perfused with saline containing eserine to prevent destruction of acetylcholine by cholinesterase the perfusate during stimulation of the nerve can be shown to contain acetylcholine. By injecting acetylcholine into the artery supplying a muscle a contraction is produced (Fig. 42, 19); this phenomenon unlike other effects of acetylcholine is remarkable in not being abolished by atropine. The response is not a twitch but a short tetanus, because the injected acetylcholine cannot reach all the muscle cells simultaneously. Skeletal muscle contains cholinesterase which is in greatest concentration near the motor end-plates, and it is therefore likely that nerve impulses, when they reach the neuromuscular junction (the end-plate), liberate acetylcholine which is quickly destroyed by cholinesterase so that a very short contraction, that is a twitch, occurs. As in other places in the body eserine, by its anticholinesterase action, enhances the effect of acetylcholine and a tetanus occurs instead of a twitch.

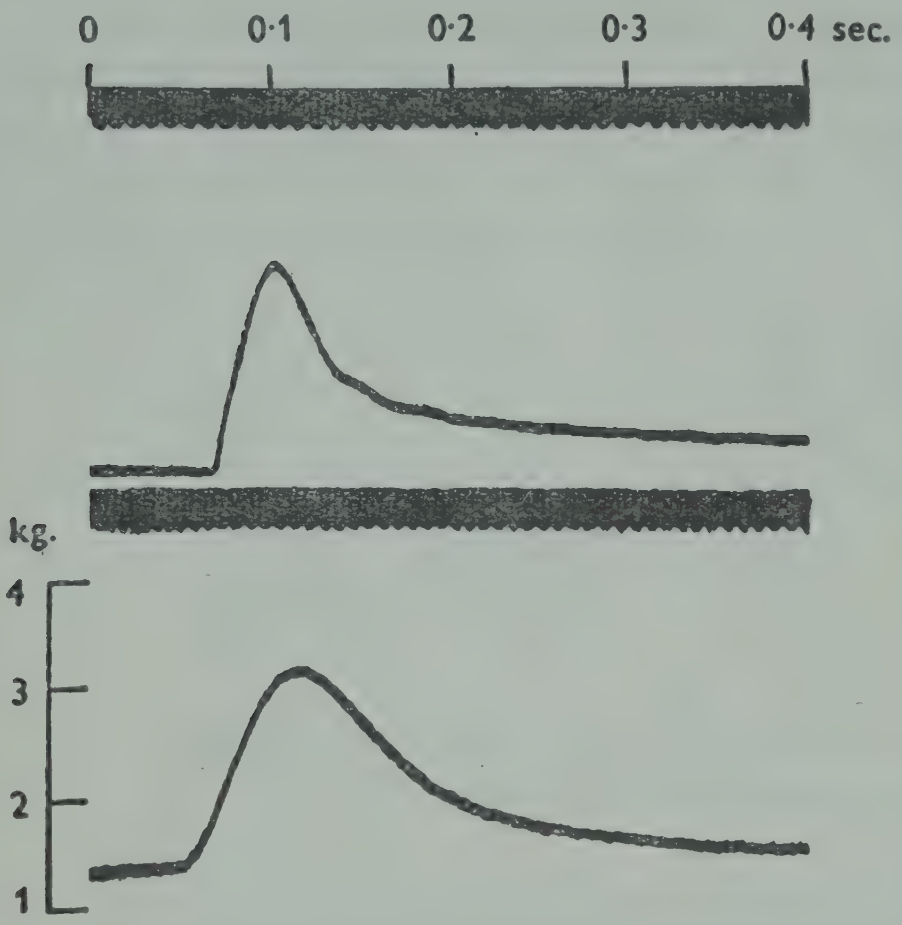


FIG. 42, 19

Reactions of the normal mammalian muscle to acetylcholine. Upper record: maximal motor nerve twitch. Lower record: response (a very short tetanus) to *close* arterial injection, during temporary arrest of the circulation to the muscle, of 2·5 μg. of acetylcholine in 0·25 ml. of saline. For close arterial injection a cannula is placed in the main artery of the limb near the muscle and branches to all muscles but the one under investigation are tied. Time in 10 msec (G. L. Brown, H. H. Dale & W. Feldberg (1936), *J. Physiol.* **87**, 402).

The theory of chemical transmission has sometimes been criticized on the basis of an apparent discrepancy; the perfusate of a muscle thrown into activity by the stimulation of its nerve contains very little ACh compared with the large amount which has to be applied externally to produce a contraction. Acetylcholine can be applied at close range through a micropipette of 0·5μ tip diameter near the end-plate of a frog muscle. By applying a positive electric pulse to the fluid in the pipette it is found that a discharge of less than 10^{-15} mole of ACh is sufficient to depolarize the end-plate and set up a muscle impulse. Since similar amounts of ACh discharged into the interior of the muscle fibre at the myoneural junction are ineffective it seems that specific ACh receptors exist on the outer surface of the end-plate membrane, that is the surface facing the motor nerve ending, and not on the side facing the myoplasm. Since an even smaller quantity of ACh would be effective at closer range, that is if released from the nerve endings lying on the end-plate, the discrepancy disappears.

When an impulse passes down a nerve, sodium enters the fibre and potassium subsequently leaves it in equivalent amount (see p. 738). When this process reaches the motor end-plate acetylcholine is liberated and locally short-circuits the muscle membrane, setting up another propagated sodium and potassium exchange along the muscle which gives rise to the muscle action potential which in turn elicits a muscular contraction. In other words the acetylcholine acts on the end-plate to produce an end-plate potential which is followed immediately by the muscle spike (action potential). By the use of intracellular electrodes Fatt and Katz have measured the quantity of electricity transferred across the end-plate membrane following on the arrival of an impulse along the motor nerve. This quantity can exceed the total ionic charge inside the terminations of the motor nerve endings. The transfer, therefore, cannot be accomplished directly by movement of ions produced by the nerve endings. It is best explained on the assumption that small quantities of acetylcholine are released from the endings and make the end-plate permeable to the free ions on either side of the membrane. This short circuit produces a high local current density. The short circuit leads to a simple depolarization and not a reversal of the membrane potential as occurs during the spike in nerve or muscle fibres; the short circuit further provides for the amplification of ionic currents which must occur at the site of transference of the impulse from minute nerve endings to the enormously greater surface of the muscle fibre.

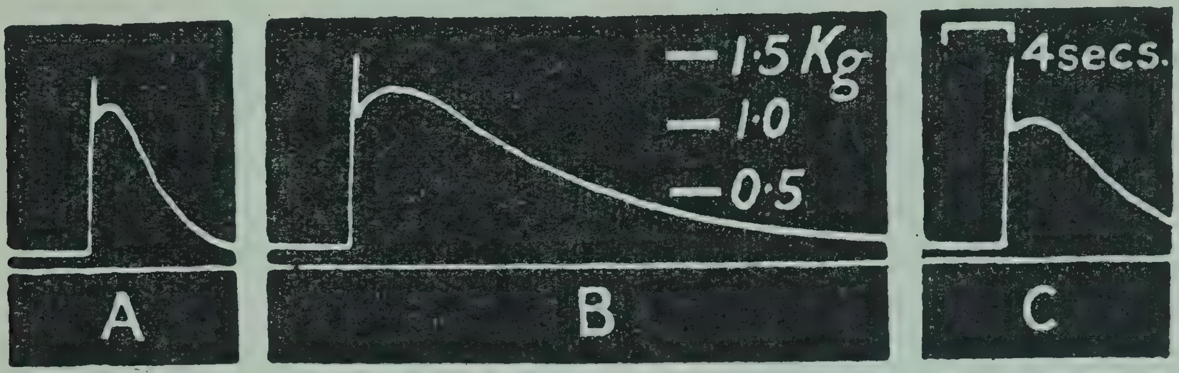


Fig. 42, 20

Cat, decerebrate. Contractions of gastrocnemius, eight days after nerve section, in response to three close arterial injections of 0·5 μg. ACh. Between A and B arterial injection of 80 μg. eserine. C, 30 min after B. Note that these records of denervated muscle were made on a slowly moving kymograph. Compare the duration of curve A with that of the lower record in Fig. 42, 19 (G. L. Brown (1937). Acetylcholine and denervated muscle. *J. Physiol.* **89,** 446).

The amount of acetylcholine released at motor nerve endings is reduced in the presence of excess magnesium ions; the end-plate potential is reduced and a block may be produced. On the other hand excess calcium in the fluid bathing the muscle increases the liberation of acetylcholine. The actual amount liberated at the ending may be a function of the relative concentrations of these two ions.

When a muscle and its nerve are at rest an electrode placed near an end-plate detects randomly recurring *miniature end-plate potentials* (m.e.p.p.) each of which is about one hundredth of the end-plate response to a nerve impulse. The m.e.p.p.s arise from the intermittent release of small quantities of acetylcholine at the nerve terminals, occasionally random discharges at several terminals may coincide and, if the muscle is treated with an anticholinesterase, may cause individual fibres to contract in a random fashion (fibrillation).

Tissues which have adrenergic nerves become more sensitive to adrenaline after denervation (p. 811). Similarly, structures deprived of their cholinergic nerves become more sensitive to acetylcholine. This is true of skeletal muscle, since after denervation a small dose of acetylcholine injected intra-arterially produces a prolonged contraction (Fig. 42, 20).

Curare, the South American arrow poison, has a remarkable effect on neuromuscular transmission which was first demonstrated by Claude Bernard in the following way. A ligature is tied firmly round the thigh of a frog to enclose all tissues except the sciatic nerve. Curare is then injected into the dorsal lymph sac. Stimulation of the sciatic nerve on the ligatured side produces a contraction of the gastrocnemius muscle which, being deprived of its circulation, cannot be reached by curare; stimulation of the sciatic nerve on the unligatured side has no effect, although direct stimulation of the gastrocnemius muscle on this side produces a contraction. This experiment shows that neither the nerve nor the muscle is affected by curare and strongly suggests that the paralysis is the result of a disturbance at the neuromuscular junction. It is now known that acetylcholine is produced as usual after curarization when a nerve impulse reaches the end-plate but it is apparently unable to depolarize the end-plate with the result that the end-plate potential is too small to excite the muscle. Curare appears to compete with acetylcholine for the receptor substance at the end-plate. The administration of curare causes motor paralysis without loss of sensation or consciousness; it causes death by asphyxia consequent on paralysis of the diaphragm. Many substances have a curare-like effect in varying degrees and are therefore used for producing muscular relaxation under general anæsthesia.

The potent toxin of *Clostridium botulinum* produces muscular paralysis by preventing the release of acetylcholine at the nerve endings, while tetanus toxin causes a persistent contraction of muscle because the nerve endings, instead of storing up acetylcholine until the arrival of nerve impulses, become leaky and allow acetylcholine to escape to provide a continuous supply to the muscle end-plates. The contractions of tetanus are relieved by curare. Decamethonium produces a persisting depolarization of the end-plate which spreads to adjacent parts of the muscle; the muscle is then insensitive to a nerve impulse.

The physical properties of the muscle membrane have already been described (p. 737).

Heat Production in Muscle.—The heat produced in frog muscle has been measured electrically by A. V. Hill by means of a thermopile and galvanometer. The thermopile with the muscle lying on it is placed in a test-tube which is immersed in a water-bath to prevent interfering fluctuations of temperature. Stimulating electrodes are connected to the two ends of the muscle. In the original experiments described in the following paragraph it was necessary to use a series of stimuli to produce enough heat to obtain a satisfactory record for analysis. It is only recently that accurate analysis of the heat produced in a single twitch has become possible.

When oxygen is excluded muscle produces heat in two phases, *initial heat* at the time of contraction and *anaerobic delayed heat* which is evolved during relaxation and afterwards. In an anaerobic contraction lactic acid appears. If now oxygen is admitted this disappears and more heat is evolved, called *aerobic delayed heat.* The amount of this heat is not nearly as much as would be expected if the lactic acid all disappeared by oxidation. As mentioned on p. 777 only about one-fifth of the lactic acid is oxidized, the remainder being converted back into glycogen. This last process is endothermic, that is, it requires energy. The oxidation of lactic and pyruvic acid provides the energy for this reaction so that the aerobic delayed heat is the heat of oxidation of these substances less the amount required for the endothermic restitution of glycogen.

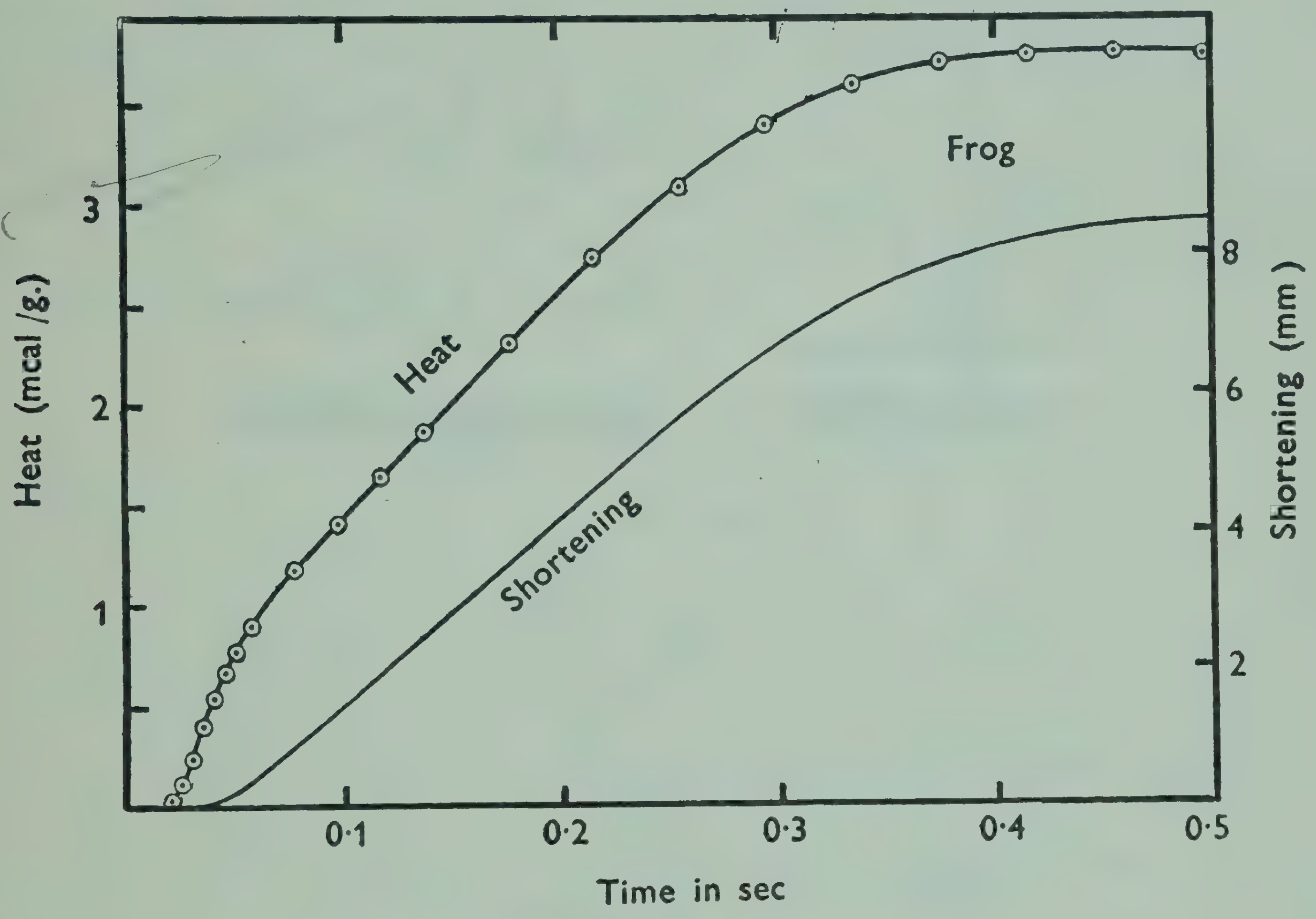


FIG. 42, 21

Heat and shortening simultaneously recorded in the isotonic twitch of a pair of frog's sartorii (120 mg., 31 mm) at 0° C and 4·3 g. load. The heat of activation starts off suddenly about 10 msec after the stimulating shock, which lasts less than 1 msec (A. V. Hill (1949), *Proc. Roy. Soc.* B, **136,** 251, Fig. 6).

Heat production may, therefore, be summarized as follows. The initial heat produced while the muscle contracts and relaxes is independent of the presence of oxygen and is associated with the breakdown of adenosine triphosphate and creatine phosphate. Following relaxation in the absence of oxygen a small amount of anaerobic delayed heat appears, which is associated with the production of lactic acid from glycogen and is less than might be expected, since at this time some energy is being used up in the resynthesis of creatine phosphate. In the presence of oxygen the aerobic delayed heat which appears over a period of several minutes following relaxation is associated with the disposal of the lactic acid by the processes described on p. 313. It should be noted that most of the heat evolved appears after the work is done. The heat is not, therefore, the source of the energy for muscular contraction. Since we do not know all the chemical processes involved in muscular contraction, any attempt to draw up a balance sheet of chemical and thermal changes is of doubtful value.

Recently improvements in myothermic technique have allowed A. V. Hill (1949) to analyse the heat produced in a single twitch of a frog or toad muscle and a summary of his results is now given. The relationship between these physical findings and the chemical events has yet to be worked out.

The heat production in a single muscle twitch occurs as (1) the *heat of activation*, an independent constant A, and (2) the *heat of shortening* (ax), the amount depending on the amount of shortening x and a constant a.

The heat of activation begins at its maximum rate very soon after the stimulus, even before shortening is detectable, and diminishes in rate as the contraction proceeds. It would occur even if shortening could be completely obviated. The heat of shortening, on the other hand, occurs at the same time as the shortening and is inseparable from (or is dependent on) the latter. Fig. 42, 21, shows the time relation between the shortening of the muscle and the heat produced after allowing for the time of propagation of the contraction over the muscle. The early, rapidly rising part of the heat curve is due to the heat of activation; the heat of shortening begins later.

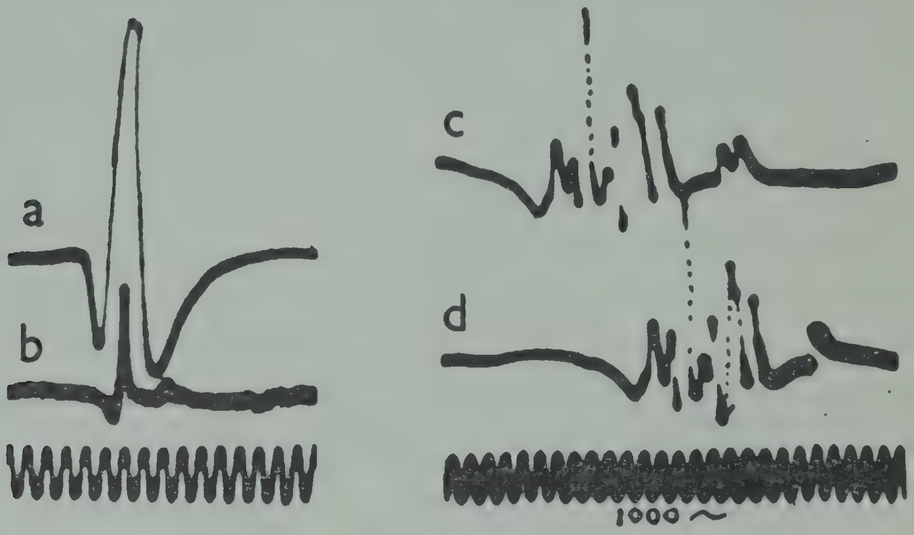


FIG. 42, 22

Human electromyograms.
a, Motor unit action potential from normal muscle; amplitude 2 mV. *b*, Single fibre potential from denervated muscle; amplitude 0·1 mV. *c* and *d*, Polyphasic motor unit action potential during attempted voluntary contraction (no visible movement) of the extensor pollicis longus muscle during early re-innervation following complete interruption and suture of the radial nerve (H. Jasper & G. Ballem (1949), *J. Neurophysiol.* **12,** 238–9).

No heat is evolved in relaxation apart from that due to the degradation into heat of the mechanical energy developed during the previous contraction phase. The relaxation heat is approximately equal to the work W dissipated. If during the time in which relaxation usually occurs the tension in the muscle is zero there is no measurable heat production.

Thus the total energy E set free in a twitch is given by

$$E = A + W + ax$$

In an isotonic contraction $W = Px$ where P is the load and the equation becomes $E = A + (P + a)x$.

Since no heat absorption could be detected even with the quickly responding apparatus used in these experiments there is no evidence for the occurrence of endothermic processes during contraction or relaxation. If endothermic processes occur in muscle contraction they are obscured by exothermic ones.*

Electromyography.—Electromyography is of value in the diagnosis of various neuromuscular disorders in which the frequency or rhythm of the nerve impulses is affected. Such conditions include peripheral nerve injuries, irritative lesions of the nerve roots, and neuromuscular disorders such as the myotonias and myasthenia gravis. The electrical activity of a muscle is best investigated by inserting very fine needle electrodes into it and recording the potentials by an amplifier and cathode-ray oscillograph. In a normal muscle a voluntary contraction produces a series of responses (see Fig. 42, 14) from motor units of the order of 1 millivolt each of which lasts about 5 msec (Fig. 42, 22*a*). Three weeks after denervation a muscle shows fibrillation due to the spontaneous activity of single muscle cells; these are rhythmically repeated potentials of the order of 0·05 mV which last only 1 msec or so. Fibrillation is increased by subcutaneous injection of prostigmine, an anticholinesterase (*cf.* p. 790); it is not abolished by anæsthesia. Fibrillation is not seen if the nerve is blocked by ischæmia or pressure or when the paralysis is due to interruption of the upper motor neurone. When re-innervation occurs fibrillation disappears and low voltage motor unit potentials appear and gradually increase as the muscle recovers (Fig. 42, 22, *c* and *d*).

---

* A. V. Hill (1949), *Proc. Roy. Soc.* B, **136,** 195.

MUSCULAR EXERCISE

Circulatory and Respiratory Adjustments.—Muscular exercise, especially if severe, involves readjustments of circulation and respiration which are described in other sections of this book. Even before exercise begins, that is before the circulatory and respiratory reflexes can be initiated, the heart quickens (Fig. 42, 23) and breathing accelerates. This anticipatory response is presumably due to the activity of the cerebral cortex and has the obvious advantage that the body is made ready to meet the additional demands of the muscles as soon as the exercise begins. The blood flow through the muscle of the arm at rest as determined by the venous occlusion plethysmograph (p. 502) is about 2 ml. per 100 g. of muscle per minute. After a short period of exercise the flow may increase to as much as 30 ml. per 100 g. per minute. If this rate of flow were maintained in all the skeletal muscles (40 kg.), then the total flow required would be 300 × 40 ml., or 12 litres per minute. The need for increased cardiac output is thus apparent. It has been shown that the greatly augmented blood flow is momentarily interrupted during each contraction because the pressure within the muscle exceeds that within the blood vessels lying around the muscle cells. While the active muscles show a vasodilatation there is a vasoconstriction in other parts of the body, perhaps mainly in the splanchnic area. The size of the vascular bed in the muscles is so much increased that the total peripheral resistance is reduced and the blood pressure does not rise as much as might be expected from the raised cardiac output (p. 518).

The main factors, in addition to those mediated through the higher centres, leading to an increase of cardiac output are as follows. The contractions and relaxations of the muscles (muscular pump, respiratory pump, p. 436) bring more blood back to the right heart and tend to cause a rise of venous pressure which, partly through the right atrial reflex (p. 466) but mainly by better filling of the ventricles in diastole (Starling's law, p. 446), causes an increase in heart rate and in the output per beat. The increased cardiac output leads to a rise of systolic blood pressure, but there is generally little alteration in the diastolic pressure. The blood is usually so well oxygenated in exercise (p. 584) that oxygen lack is only rarely a factor in producing acceleration of the heart, but the production of carbon dioxide is much increased and this, either by direct action on the cardiac centres in the medulla or reflexly via the chemoreceptors of the carotid body, contributes to the increase in heart rate.

Respiration is quickened, as already mentioned, before exercise begins and during exercise it continues to be stimulated, not by oxygen lack but mainly by the increased production of carbon dioxide. The hastening and deepening of respiration, especially in severe exercise, cannot be fully accounted for by the rise in pCO_2 and fall of pH of the blood. The alveolar pCO_2 can be raised artificially at rest by inhalation of air containing carbon dioxide (p. 555) but the pulmonary ventilation achieved at a given alveolar pCO_2 in this way is much less than in exercise which produces the same rise in alveolar pCO_2. In very severe exercise anoxia may be an additional factor but more important is the lowered threshold of the respiratory centre to the pCO_2 of the arterial blood. In the recovery period immediately after exercise the alveolar pCO_2 level falls below the pre-exercise level and administration of carbon dioxide in the inspired air causes a greater increase of pulmonary ventilation than it does at rest. The lowering of the threshold of the respiratory centre can be attributed to the increased body temperature and the raised concentration of blood lactate. Some experiments suggest that in the muscles, probably in their blood vessels, there are nerve endings, which, on stimulation by metabolites, produce reflexly an increase in respiration. It is likely, however, that the main factor producing the increase in respiration in exercise is a lowering of the threshold of the respiratory centre as just explained.

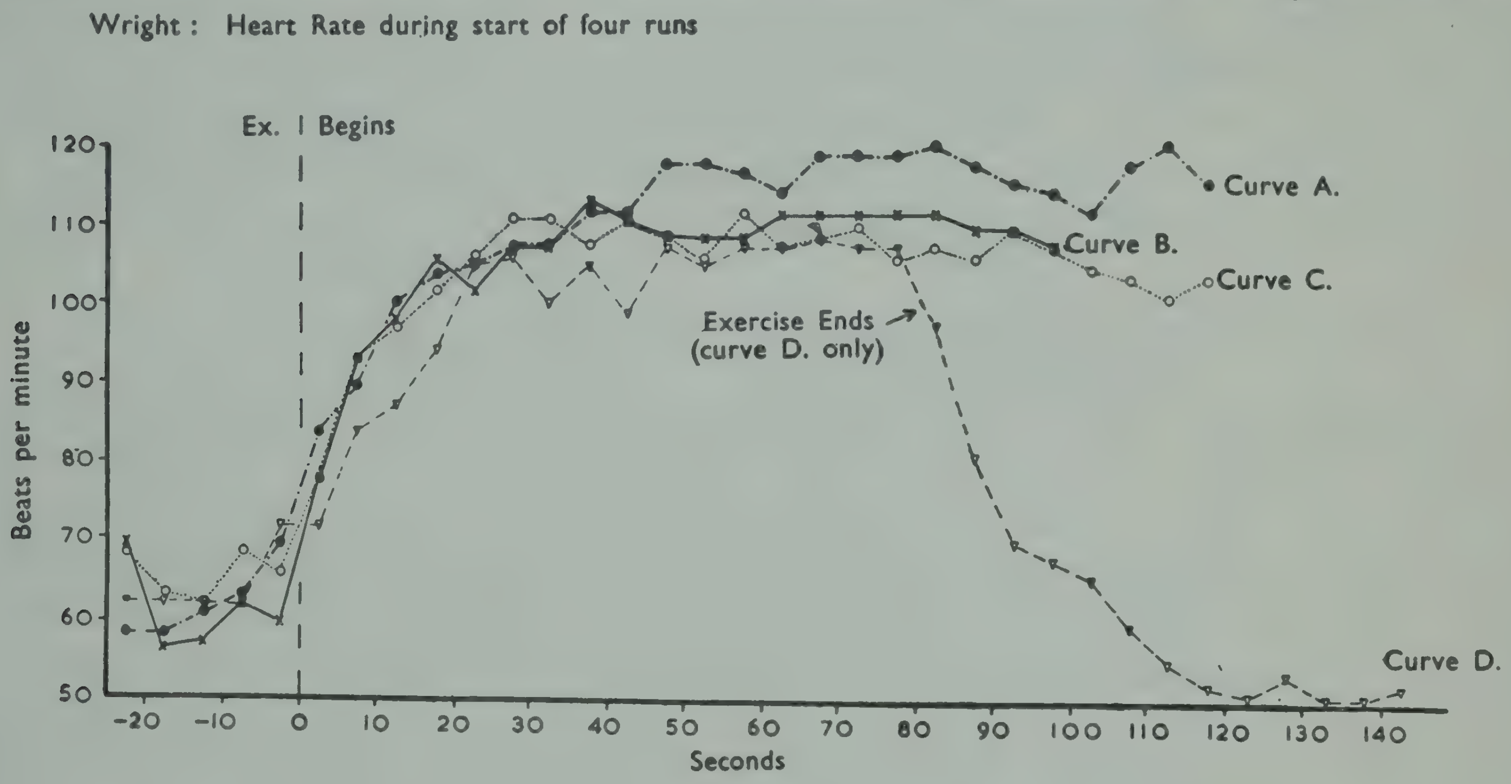


Fig. 42, 23

Subject, D. M. Wright (Marathon runner). Graphs of heart rate in five-second intervals at the start of various runs on the treadmill at 6 m.p.h. Curve A, start of 5·9 mile run. Curves B, C and D, three separate starts at half-hour intervals taken one year after Curve A.

(*By courtesy of J. A. C. Knox.*)

The exercising muscle is not only more liberally supplied with blood but the low oxygen tension in the cells, the increased formation of carbon dioxide and acid metabolites all favour an increased uptake of oxygen from the blood. More oxygen is taken out of each ml. of blood

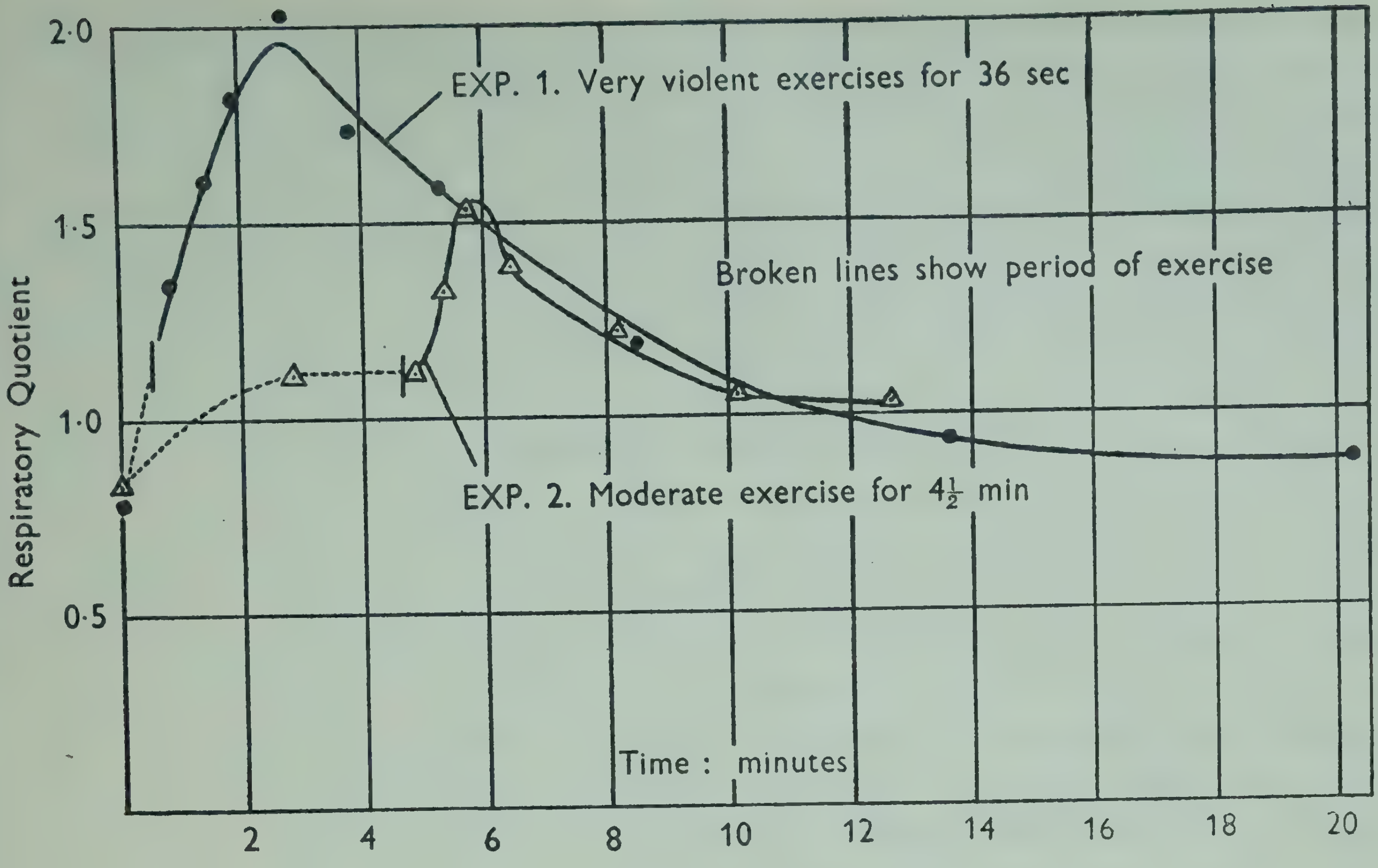


FIG. 42, 24

Subject, C.N.H.L. The respiratory quotient during and after muscular exercise. These figures show the initial phase of recovery only. The final phase is shown in Fig. 42, 25 (A. V. Hill, C. N. H. Long & H. Lupton (1924), *Proc. Roy. Soc.* B, London, **96,** 468).

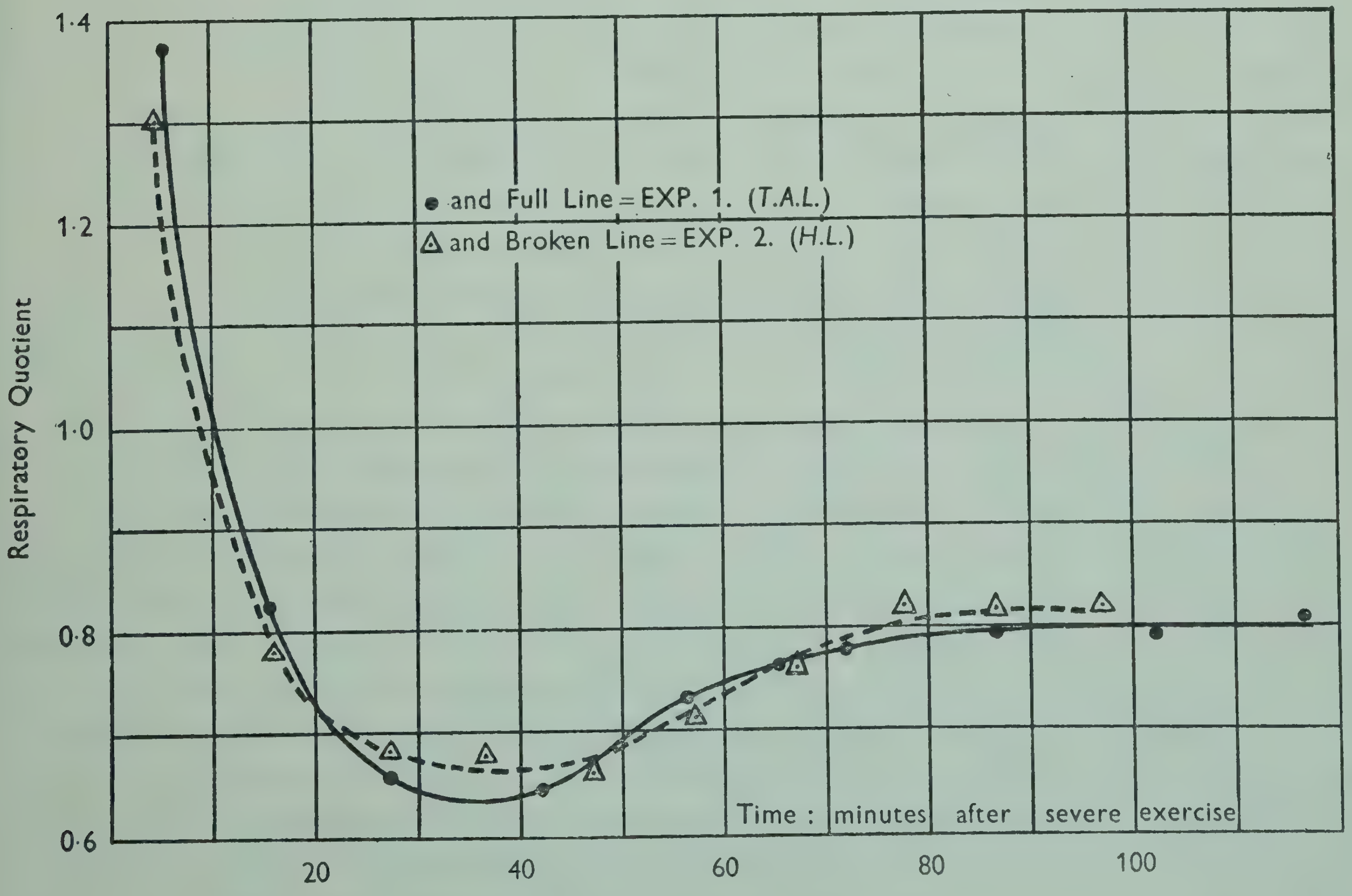


FIG. 42, 25

The respiratory quotient after severe muscular exercise. Note that in the later phases, while carbon dioxide is being retained to compensate for that initially driven off, the respiratory quotient falls to a very low level, returning to its final value in about 80 minutes (A. V. Hill, C. N. H. Long & H. Lupton (1924), *Proc. Roy. Soc.* B, **96,** 472).

so that the arteriovenous oxygen difference is greater in exercise than at rest. Although by the simultaneous operation of all these factors large increases in oxygen requirements can be met, an oxygen debt (Fig. 42, 27 and 28) is always incurred in exercise unless it is very mild.

The Chemical Changes in Exercise.—It is explained on p. 154 how the respiratory quotient is related to the kind of food which is being burnt. The R.Q. at rest is about 0·85 on the average and vigorous exercise causes a rise. This suggests that a slightly larger proportion of carbohydrate is being oxidized. It might be supposed that a clue to the type of fuel used could be derived from the R.Q. of the metabolism in excess of the resting value, that is, the ratio of excess carbon dioxide production over the resting value to the excess of oxygen consumption over the resting value. The R.Q. of the excess metabolism for a complete period of exercise and recovery usually lies between 0·85 and 1·0. This suggests that much the same fuel is burnt during exercise as at rest. If, however, the R.Q. is measured over short periods during and after exercise apparently anomalous results may be obtained. At the beginning of exercise carbon dioxide is blown off in excess, partly because of the stimulation of respiration by the higher centres and partly because, in strenuous exercise, lactic acid produced by muscular metabolism is liberated into the blood, where it is buffered by the bicarbonate (p. 644) with formation of carbonic acid. The latter stimulates respiration and more carbon dioxide is got rid of in the lungs. Since oxygen is involved in neither of these processes the ratio $\frac{\text{CO}_2\text{ produced}}{\text{O}_2\text{ used}}$ becomes larger and may be over 1 (Fig. 42, 24). After vigorous exercise has ended the oxygen intake remains high because lactic acid and other substances which had accumulated in the muscles must be oxidized. With the oxidation of lactate, base is set free for taking up a further supply of carbon dioxide. During this process less carbon dioxide is expelled from the blood via the lungs and so the R.Q. may fall below 0·7 (Fig. 42, 25). The increased energy output during exercise inevitably gives rise to an increased production of heat since only some 25 per cent. of the energy liberated is converted into mechanical work. The raised blood temperature results in cutaneous vasodilatation (p. 206) and an increased heat loss. In strenuous exercise the augmented heat loss may be incapable of balancing the increased production of heat and the body temperature may rise several degrees centigrade during the exercise (Fig. 42, 26).

Several writers in describing events during muscular exercise have used the *submarine analogy*. The submarine on the surface where oxygen is available charges its accumulators and thus stores energy which can be called on when the submarine is submerged and oxygen is not available. The accumulators can discharge under anaerobic conditions but they can be recharged only when oxygen is available for the internal combustion engines, that is under aerobic conditions. Although this analogy cannot bear detailed comparison with events occurring in muscle, it helps one to formulate a clearer idea of reactions occurring there. From the description already given it is evident that three chemical reactions can produce heat anaerobically:

(1) adenosine triphosphate (ATP) → adenosine diphosphate (ADP) + phosphoric acid;

(2) creatine phosphate → creatine + phosphoric acid;
(3) glycogen → pyruvic acid and lactic acid.

While the anaerobic reactions are occurring it can be supposed that the muscle accumulators are discharging. In a short burst of severe exercise, say a hundred yards' sprint lasting 10 to 12 seconds, the volume of oxygen that can be taken in is very small compared with the volume actually required and energy must, therefore, be derived from anaerobic processes.

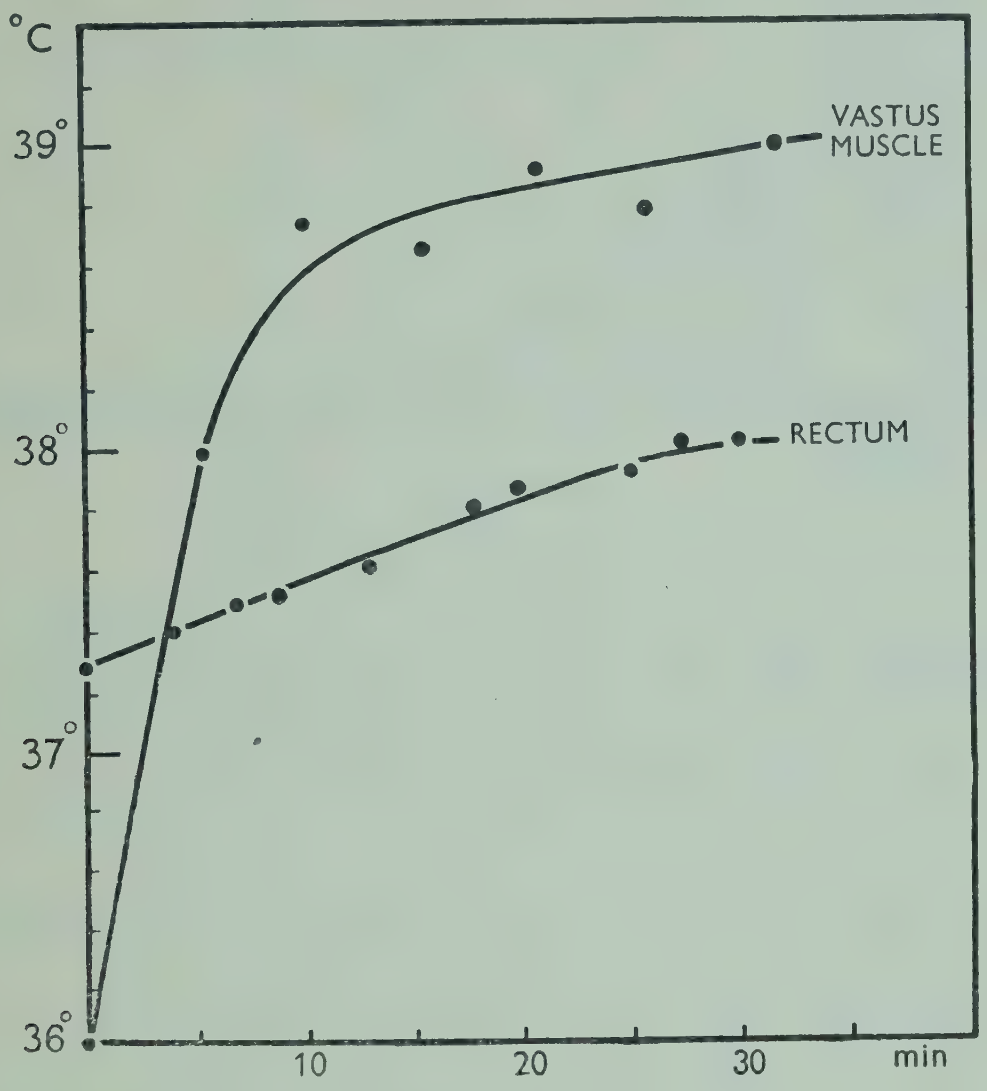


FIG. 42, 26

Temperature measured in the lateral vastus muscle and in the rectum during a work of 660 kg.m per min (E. Asmussen and O. Boje (1950), *J. phys. Educ.* **42,** 48).

This results in an oxygen debt which is paid back through deep and frequent breathing that continues long after the exercise is over. The additional amount of oxygen consumed above the resting oxygen consumption reveals the amount of the oxygen debt (Fig. 42, 27). During this period the accumulators are being charged, or, in other words, the three reactions listed above are proceeding in the reverse direction as the result of energy supplied by oxidation of lactic acid and other materials. In mild exercise the anaerobic mechanism may be called on only at the beginning (Fig. 42, 28) because as exercise proceeds the circulatory and respiratory adjustments allow as much oxygen to be taken in as is required and a steady state is therefore reached although the oxygen debt is still unpaid. The initial oxygen debt is paid off at the end of the exercise. Even a small initial oxygen debt is apparently not paid off during the period of exercise. The body does not seem able to take in more oxygen than is required for current expenditure and it often fails to do even that. The lactic acid level in the blood during mild exercise such as walking does not rise above the resting value of 15 mg. per 100 ml. This is interpreted as meaning that

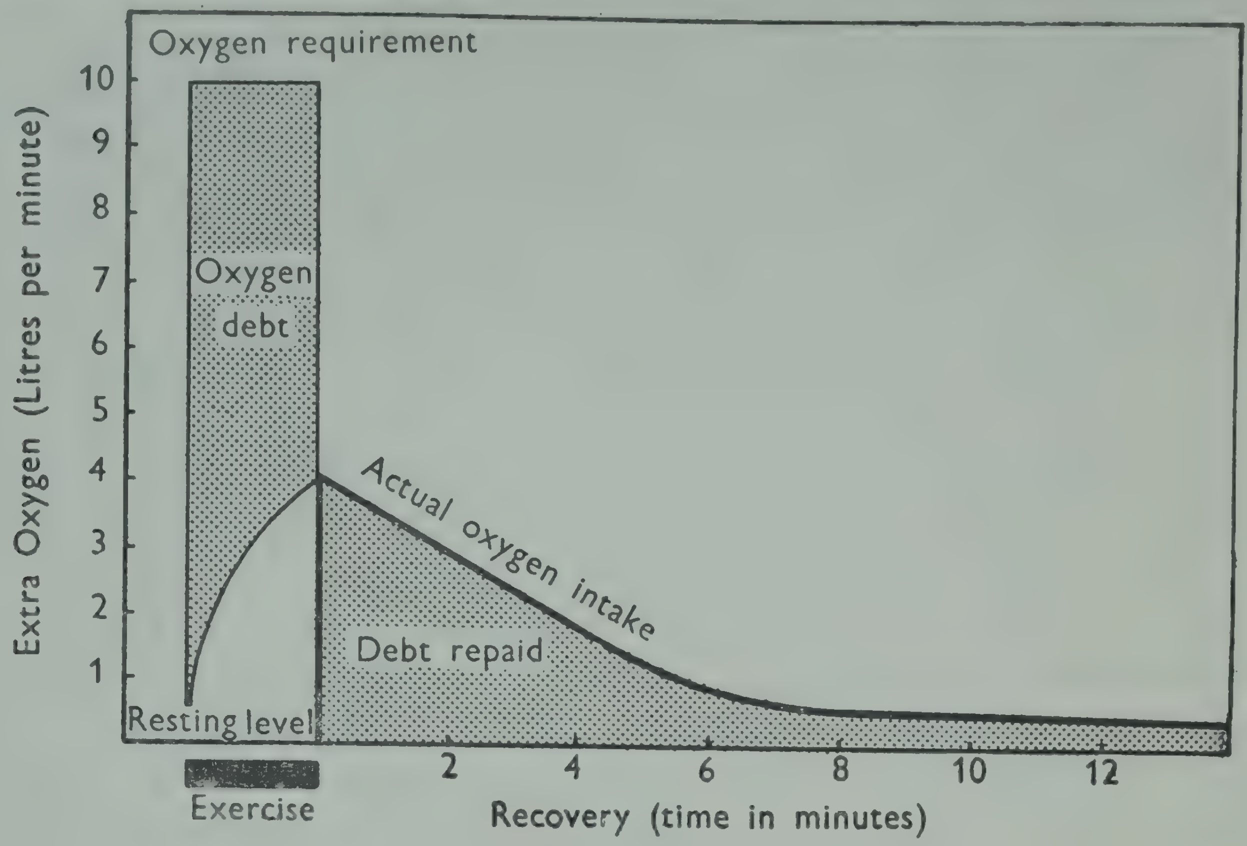


FIG. 42, 27

Diagram illustrating the relationship between 'oxygen requirement' and 'oxygen intake' in severe exercise. The line marked 'Actual oxygen intake' has been drawn from experimental findings. The column marked 'Oxygen requirement' is constructed by making the shaded area marked 'Oxygen debt' equal to the shaded area marked 'Debt repaid.' (*After* M. G. Eggleton from K. Furusawa, A. V. Hill, C. N. H. Long & H. Lupton (1924), *Proc. Roy. Soc.* B, **97**, 155.)

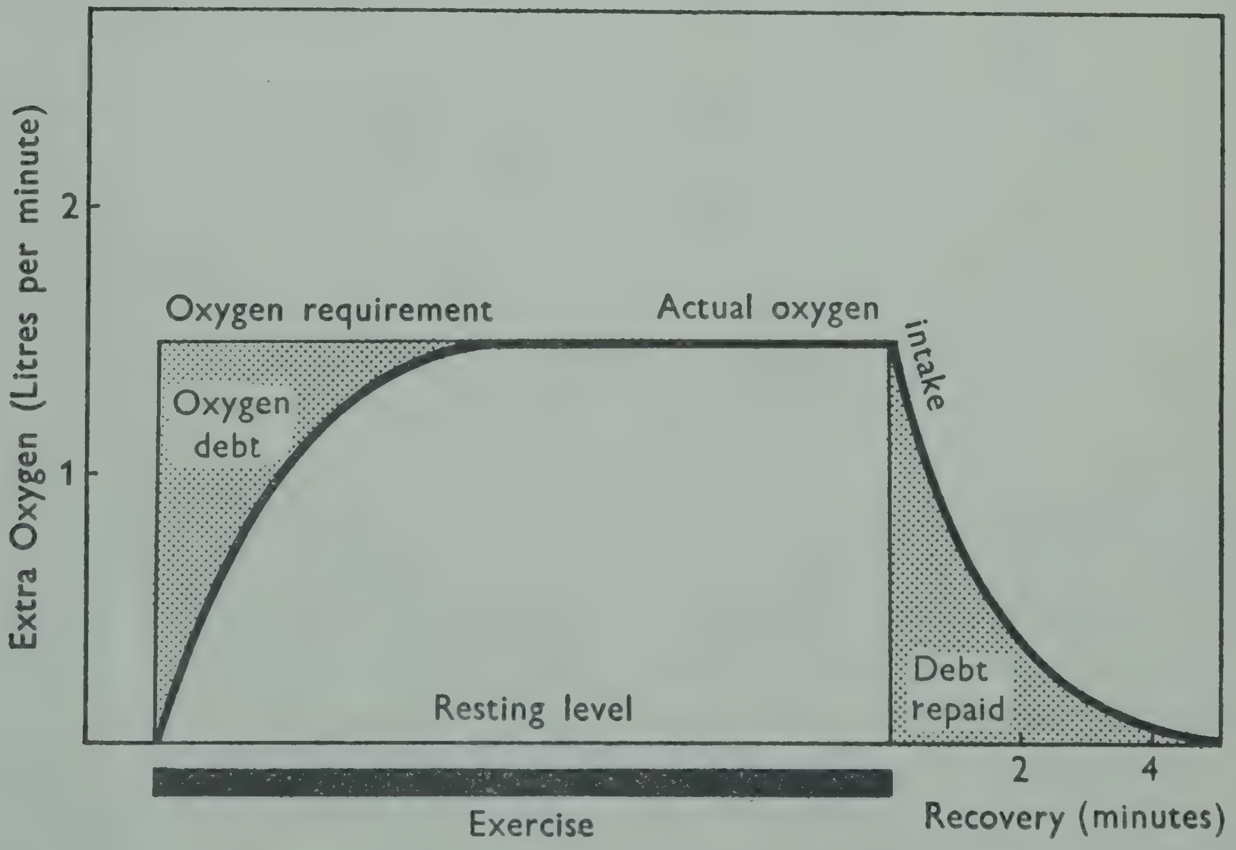


FIG. 42, 28

Diagram illustrating the relationship between 'oxygen requirement' and 'oxygen intake' in mild exercise. The method of constructing this diagram is given in the legend to Fig. 42, 27. (*After* M. G. Eggleton from K. Furusawa, A. V. Hill, C. N. H. Long & H. Lupton (1924), *Proc. Roy. Soc.* B, **97**, 155.)

lactic acid is disposed of as fast as it is formed. It does not, therefore, accumulate in the muscle tissue. In more vigorous exercise, however, lactic acid is formed so rapidly that the mechanisms for oxidizing it or reconverting it to glycogen are overloaded. It therefore appears in increased amount in the blood so that in severe exercise the blood lactate rises as high as 50 mg. per 100 ml. and sometimes up to 200 mg. per 100 ml. The appearance of lactic acid in the blood is in some way related to the state of 'training' of the subject (Fig. 42, 29). The glycogen-lactic acid mechanism is not quite so simple as may at first sight appear, since even in complete exhaustion a muscle still possesses considerable glycogen stores; thus the glycogen content of a frog muscle stimulated to exhaustion may fall only from 0·7 to 0·5 g. per cent. It is likely that the maximum oxygen debt which can be incurred is related to the maximum tolerable lactic acid production.

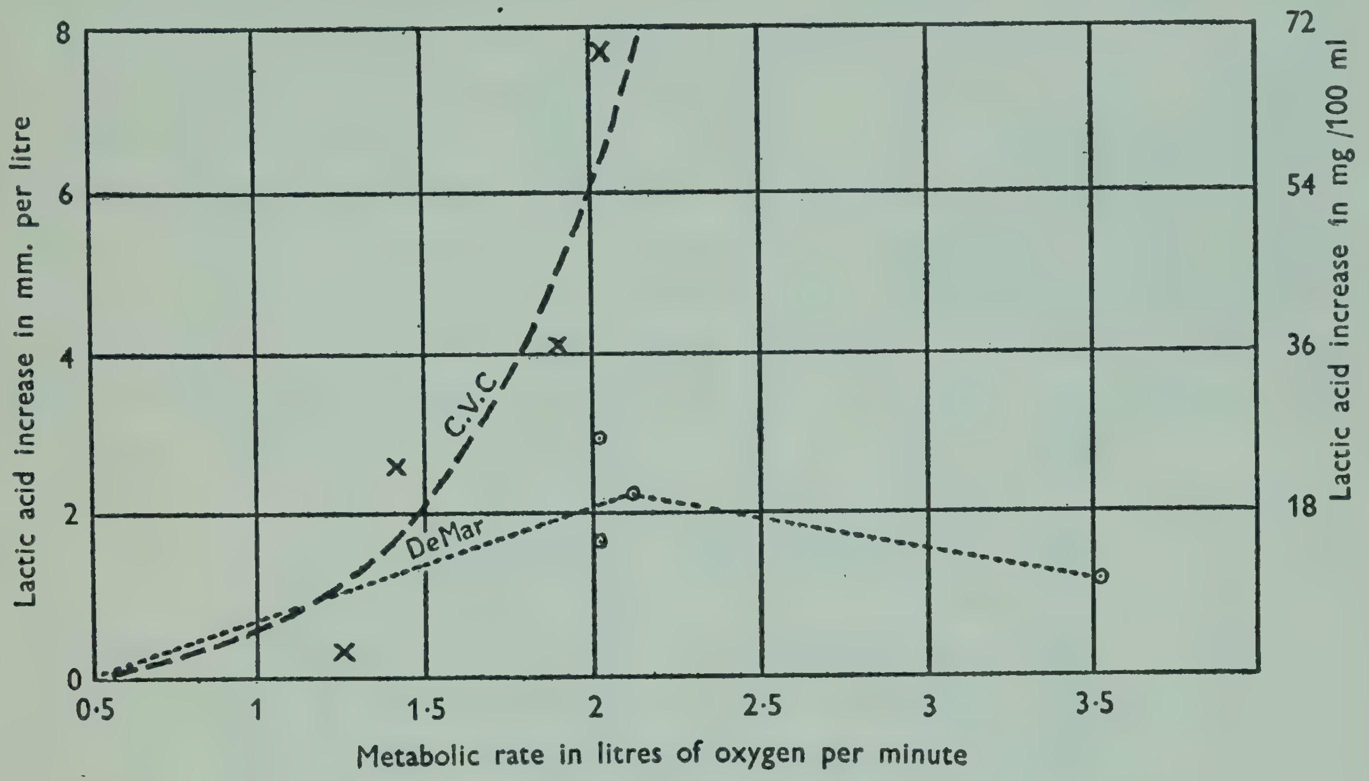


FIG. 42, 29

The increase in blood lactic acid occurring in exercise, the severity of which is indicated by the metabolic rate. De Mar was a trained Marathon runner. Even at an oxygen consumption of 3·5 l./min the resynthesis of lactic acid kept pace with the breakdown so that the blood lactic acid did not rise. The subject, C.V.C., had never had extended experience in any form of bicycle exercise. The observation on De Mar at 3·5 litres was obtained from blood drawn after a fifteen-mile run (A. V. Bock, C. Vancaulaert, D. B. Dill, A. Folling & L. M. Hurxthal (1928), Muscular activity. *J. Physiol.* **66**, 153).

The effects of exercise of varying degrees of severity are given in Table 42, 30, and Fig. 42, 31. It will be seen that the increase of oxygen consumption for any given exercise is much greater than the increase in cardiac output. For example (Table 42, 30), running at 7·5 m.p.h. involves a twelvefold increase in oxygen consumption but only a fivefold increase in cardiac output. The same phenomenon is illustrated in Fig. 42, 31. The increased amounts of oxygen can be taken up in the lungs because the blood coming to them is less saturated than usual and because a greater quantity of blood is going through the lungs in unit time. The venous saturation has been calculated (see Fick principle, p. 449) and is plotted in Fig. 42, 31. This graph also shows that within the range of the observations the pulse rate increased from 70 to 165 per minute whereas the cardiac output rose from 5 to 25 litres per minute. A simple calculation shows that the stroke volume (output per beat) must have increased from 70 ml. up to 150 ml.

TABLE 42, 30

Approximate figures for the cardiac output and oxygen consumption during exercise of varying severity

Exercise	Cardiac Output l./min	Total Oxygen Consumption of Body l./min
Rest	5	0·25
Walking 2 m.p.h. . .	10	0·8
Walking 5 m.p.h. . .	22	2·5
Running 7·5 m.p.h. . .	25	3·0
Very severe exercise (max. athletic effort) .	30	4·0

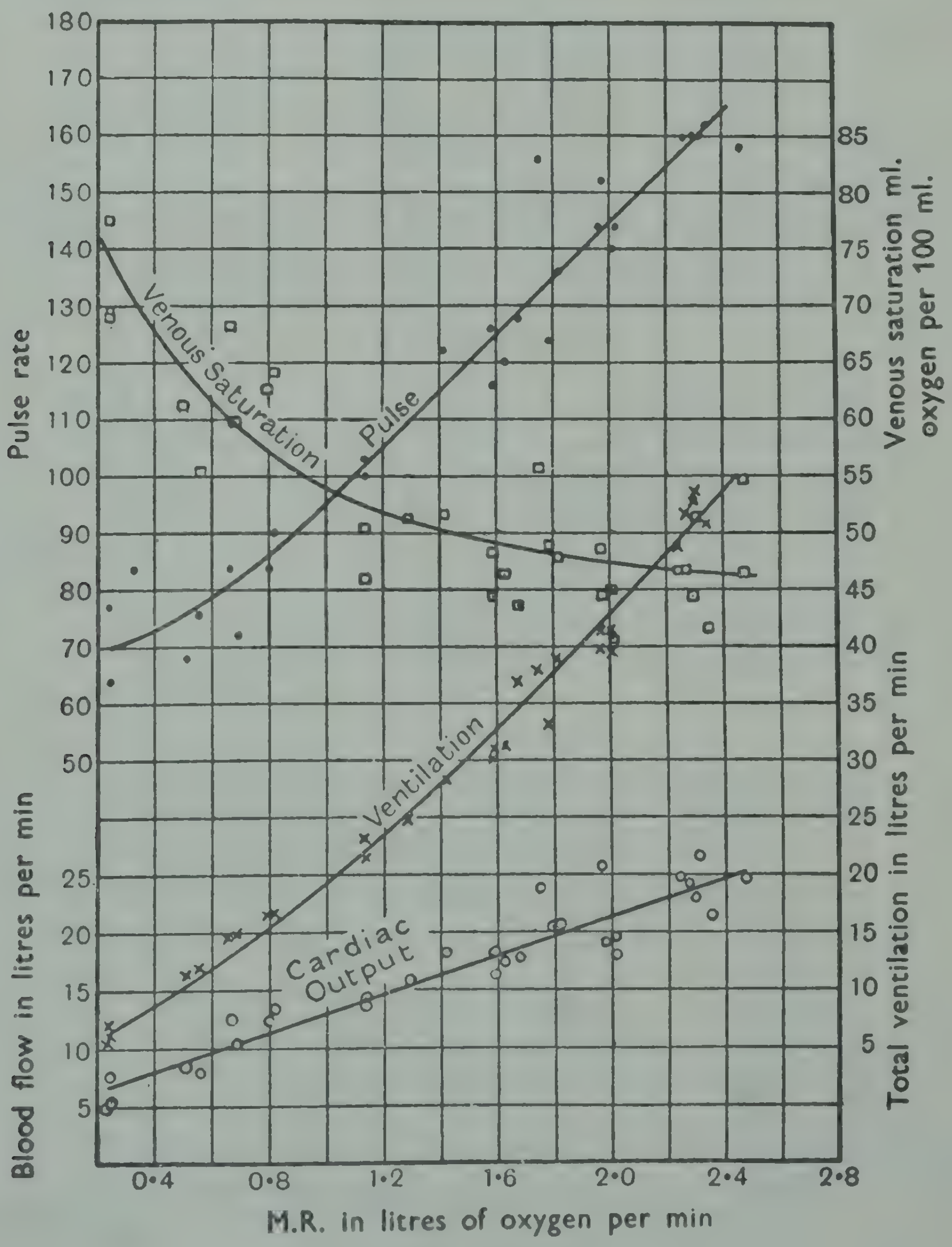


FIG. 42, 31

Data for exercise of varying severity plotted as a function of metabolic rate (M.R.) on a stationary bicycle ergometer. Periods of work: 20 minutes or longer. Subject, D.B.D. (A. V. Bock, C. Vancaulaert, D. B. Dill, A. Fölling & L. M. Hurxthal (1928). *J. Physiol.* **66,** 141).

Second Wind.—At the beginning of moderately severe exercise, say a long-distance run, the breathing becomes increasingly laboured; after some time the discomfort disappears and the subject says he has got his *second wind.* The phenomena of second wind are variable. In some people the transition from dyspnœa to hyperpnœa occurs suddenly, while in others the change is quite slow. The rate of transition has little to do with training as the phenomenon is often quite marked in the untrained person. During the dyspnœic period the alveolar carbon dioxide tension is higher than normal, but when the subject gets his second wind the alveolar carbon dioxide tension falls and there is a fall in the ventilation rate which is responsible for the feeling of comfort. At the same time sweating may occur and the body temperature may rise. The respiratory quotient usually falls. A satisfactory explanation of this phenomenon is not yet forthcoming but it is likely that the early dyspnœa of exercise is due to the liberation of lactic acid which displaces carbon dioxide from bicarbonate; this additional carbon dioxide causes stimulation of the respiratory centre. In time, when this additional carbon dioxide is eliminated by vigorous breathing and the circulatory adjustments to meet the demands of exercise are completed and less lactic acid reaches the blood, the body reaches a steady state in which respiration is less laboured.

SMOOTH MUSCLE

The muscle of the intestinal tract, the uterus, the spleen, and the blood vessels is called smooth, or visceral, muscle. Smooth muscle cells have a relatively simple structure; they are long (15 to 500 μ), narrow (2 to 20 μ) cells tapering towards the ends with a single long narrow nucleus about the middle of the cell. They may show a faint longitudinal striation but never a transverse, and there is no definite sarcolemma. Many smooth muscle cells receive both sympathetic and parasympathetic nerve fibres but there are no elaborate motor end-plates. Smooth muscle does not appear to be organized in motor units like skeletal muscle and its blood supply is relatively smaller.

The movements of smooth muscle are very much slower than those of skeletal muscle but they should not, therefore, be thought of as being much weaker. The force exerted on the child by the smooth muscle of the uterus during parturition is very large indeed. Whereas skeletal muscle is paralysed when its nerves are cut, smooth muscle continues its rhythmic activity, though perhaps in a modified way, when its autonomic connexions have been severed. Whether this ‘spontaneous’ activity is myogenic, that is to say, an intrinsic property of the muscle fibres, is difficult to decide since most areas of smooth muscle contain nerve fibres and even ganglion cells. The sympathetic and parasympathetic nerves to a given group of visceral muscle fibres have usually opposite effects one being motor and the other inhibitor, but the effect obtained on stimulation of, say, the sympathetic nerve is different in different organs and not necessarily the same in the same organ at different times. The tone of the muscle at the time of the experiment is one of the factors determining whether a motor or an inhibitory response is obtained.

The membrane potential (see p. 737) of smooth muscle cells of the colon has been measured by an intracellular electrode. It is about 60 mV on the average but in any one fibre the level fluctuates. Spontaneous activity is accompanied by a rapid depolarization (up to 8 mV) followed by a small repolarization; each depolarization shows as a spike on the record. Acetylcholine causes an increase in the tension of the muscle, a fall in membrane potential and an increase in the frequency of the spikes; adrenaline produces exactly the opposite effects.

Although the strength-duration curves (p. 758) of smooth muscle have the same general shape as those for nerve or skeletal muscle the chronaxie is quite long, in the case of the frog's stomach about 50 msec. Thus smooth muscle is relatively unresponsive to induced currents although it may readily be stimulated by the constant current from a battery (galvanic stimulation). When one part of a sheet of smooth muscle is stimulated a wave of contraction passes over the whole sheet, presumably because there is protoplasmic continuity between adjacent fibres as in cardiac muscle. It is possible, however, that the fine nerve plexuses accompanying the muscle cells play a part in conveying the stimulus for the wave of contraction.

Smooth muscle is responsive to alterations in temperature. Thus when the abdomen is opened and a loop of intestine is exposed to cool air it contracts but relaxes again if it is warmed. A fall in pH usually initiates a contraction and an increase a relaxation. Because of its spontaneous activity and the effects just described it is difficult to say what the resting length of a piece of smooth muscle is.

The chemical processes involved in the contraction of smooth muscle are probably much the same as in skeletal muscle as it contains enzyme systems similar to those in skeletal muscle. However, smooth muscle when contracted consumes no more oxygen than it does when relaxed, and for this reason it is difficult to demonstrate fatigue in visceral muscle.

Heart muscle, like voluntary muscle, can utilize both glucose and lactic acid but it can, in addition, obtain energy, in the absence of available carbohydrate, directly from the fatty acids of the blood or of the heart muscle cells (p. 455).

REFERENCES

BALDWIN, E. (1952). *Dynamic Aspects of Biochemistry.* 2nd ed. Cambridge University Press.

BANNISTER, R. G., CUNNINGHAM, D. J. C. & DOUGLAS, C. G. (1954). The carbon dioxide stimulus to breathing in severe exercise. *J. Physiol.* **125,** 90–117.

BARKER, D. (1948). The innervation of the muscle spindle. *Quart. J. micr. Sci.* **89,** 143–186.

DONALD, K. W., BISHOP, J. M., CUMMING, G. & WADE, O. L. (1955). The effect of exercise on the cardiac output and circulatory dynamics of normal subjects. *Clin. Sci.* **14,** 37–73.

DRAPER, M. H. & HODGE, A. J. (1949). Studies on muscle with the electron microscope. *Aust. J. exp. Biol. med. Sci.* **27,** 465–503.

DUBUISSON, M. (1954). *Muscular Contraction.* Oxford: Blackwell.

EGGLETON, M. G. (1936). *Muscular Exercise.* London: Kegan Paul.

FATT, P. (1954). Biophysics of junctional transmission. *Physiol. Rev.* **34,** 674–710.

GEMMILL, C. L. (1942). The fuel of muscular energy. *Physiol. Rev.* **22,** 32–53.

GRODINS, F. S. (1950). Analysis of factors concerned in regulation of breathing in exercise. *Physiol. Rev.* **30,** 220–239.

HILL, A. V. (1926). *Muscular Activity.* Baltimore: Williams, Wilkins Co.

HILL, A. V. (1932). Revolution in muscle physiology. *Physiol. Rev.* **12,** 56–67.

HILL, A. V. (1951). The mechanics of voluntary muscle. *Lancet,* ii, 947–951.

HILL, A. V. & others (1950). A discussion on muscular contraction and relaxation: their physical and chemical basis. *Proc. Roy. Soc.* B, **137,** 40–87.

HINSEY, J. C. (1934). The innervation of skeletal muscle. *Physiol. Rev.* **14,** 514–585.

HUNT, C. C. & KUFFLER, S. W. (1954). Motor innervation of skeletal muscle: multiple innervation of individual muscle fibres and motor unit function. *J. Physiol.* **126,** 293–303.

Kuffler, S. W. & Hunt, C. C. (1952). The mammalian small-nerve fibres: a system for efferent regulation of muscle spindle discharge. *Res. Publ. Ass. nerv. ment. Dis.* **30,** 24–47.

Mommaerts, W. F. H. M. (1950). *Muscular Contraction. A Topic in Molecular Physiology.* London: Interscience.

Morehouse, L. E. & Miller, A. T. (1948). *Physiology of Exercise.* St Louis: Mosby.

Perry, S. V. (1956). Relation between chemical and contractile function and structure of the skeletal muscle cell. *Physiol. Rev.* **36,** 1–76.

Schneider, E. C. & Karpovich, P. V. (1948). *Physiology of Muscular Activity.* 3rd ed. London: Saunders.

Szent-Györgyi, A. (1953). *Chemical physiology of contraction in body and heart muscle.* New York: Academic Press.

Tiegs, O. W. (1953). Innervation of voluntary muscle. *Physiol. Rev.* **33,** 90–144.

Weddell, G., Feinstein, B. & Pattle, R. E. (1944). The electrical activity of voluntary muscle in man under normal and pathological conditions. *Brain,* **67,** 178–257.

CHAPTER 43

AUTONOMIC NERVOUS SYSTEM

IN discussing the control of the heart beat, the calibre of the blood vessels, and the movements of the alimentary canal, the sympathetic and parasympathetic nerve supply was considered in each case. We shall discuss here only general features of the autonomic system; this includes fibres which run from the central nervous system to the viscera as well as afferent fibres from the viscera.

The autonomic nervous system has central connexions which are being slowly worked out. The frontal lobe of the cerebral cortex has an influence on the cardiovascular system and the intestinal tract by virtue of its connexions with the hypothalamus but the main control of the autonomic system seems to be located in the hypothalamus itself. Important reflex centres, including the vagus centres, are found in the medulla oblongata. These subjects are considered in other chapters of this book.

The autonomic fibres leave the cord in two great 'outflows,' the thoracolumbar (sympathetic) and craniosacral (parasympathetic).

- Autonomic Nervous System
  - Parasympathetic outflow
    - Cranial
      - Tectal (III)
      - Bulbar (VII, IX, X, XI)
    - Sacral
      - Pelvic splanchnic nerves
  - Sympathetic outflow
    - From all thoracic and upper two lumbar segments of cord

Unlike the somatic outflow which supplies the skeletal muscles the autonomic fibres to the viscera are interrupted by synapses. The fibre emerging from the central nervous system as a preganglionic fibre runs to synapse in a ganglion from which a second, postganglionic, fibre supplies gland or muscle cells. A preganglionic nerve impulse may, on reaching a synapse, set up impulses in as many as twenty postganglionic fibres. In the case of the sympathetic system the synapses occur in the ganglia of the sympathetic trunk or in more peripheral ganglia which lie along the anterior surface of the abdominal aorta. It has recently been shown that there are also intermediate ganglia lying along the course of the white rami communicantes and also in relation to nerves other than those of the thoracolumbar outflow. Most of their postganglionic fibres are distributed with the spinal nerves but some enter visceral branches. The parasympathetic ganglia commonly lie on or in the organ supplied. In all cases the ganglia can be located by the method used by Langley which depends on the fact that the alkaloid nicotine, after first stimulating ganglion cells, paralyses them by depolarization without having any effect on nerve fibres. Thus, if after painting a ganglion with nicotine the characteristic effect normally produced in a viscus by electrical stimulation of its autonomic nerve is abolished, this ganglion must contain synapses at which postganglionic fibres begin.

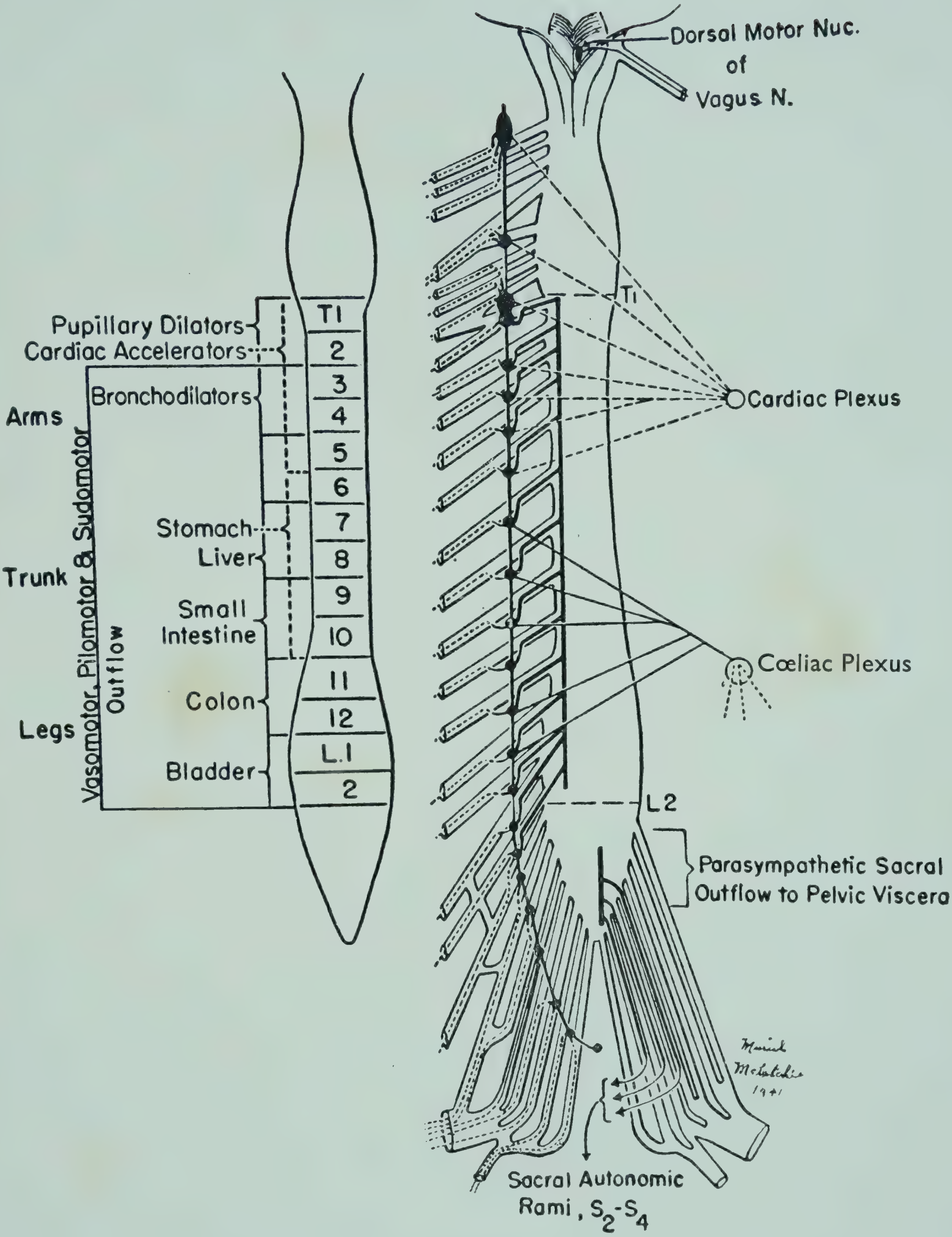

Dorsal Motor Nuc. of Vagus N.
T1
Cardiac Plexus
Cœliac Plexus
L2
Parasympathetic Sacral Outflow to Pelvic Viscera
Sacral Autonomic Rami, S_2-S_4
Muriel McLatchie 1941
Pupillary Dilators
Cardiac Accelerators
Bronchodilators
Arms
Trunk
Legs
Vasomotor, Pilomotor & Sudomotor Outflow
Stomach
Liver
Small Intestine
Colon
Bladder
T1
2
3
4
5
6
7
8
9
10
11
12
L.1
2

FIG. 43, 1

The autonomic outflow. ——— Preganglionic axon. – – – Postganglionic axon (J. C. White, R. H. Smithwick & F. A. Simeone (1952), *The Autonomic Nervous System*, 3rd ed. p. 24. London: Henry Kimpton).

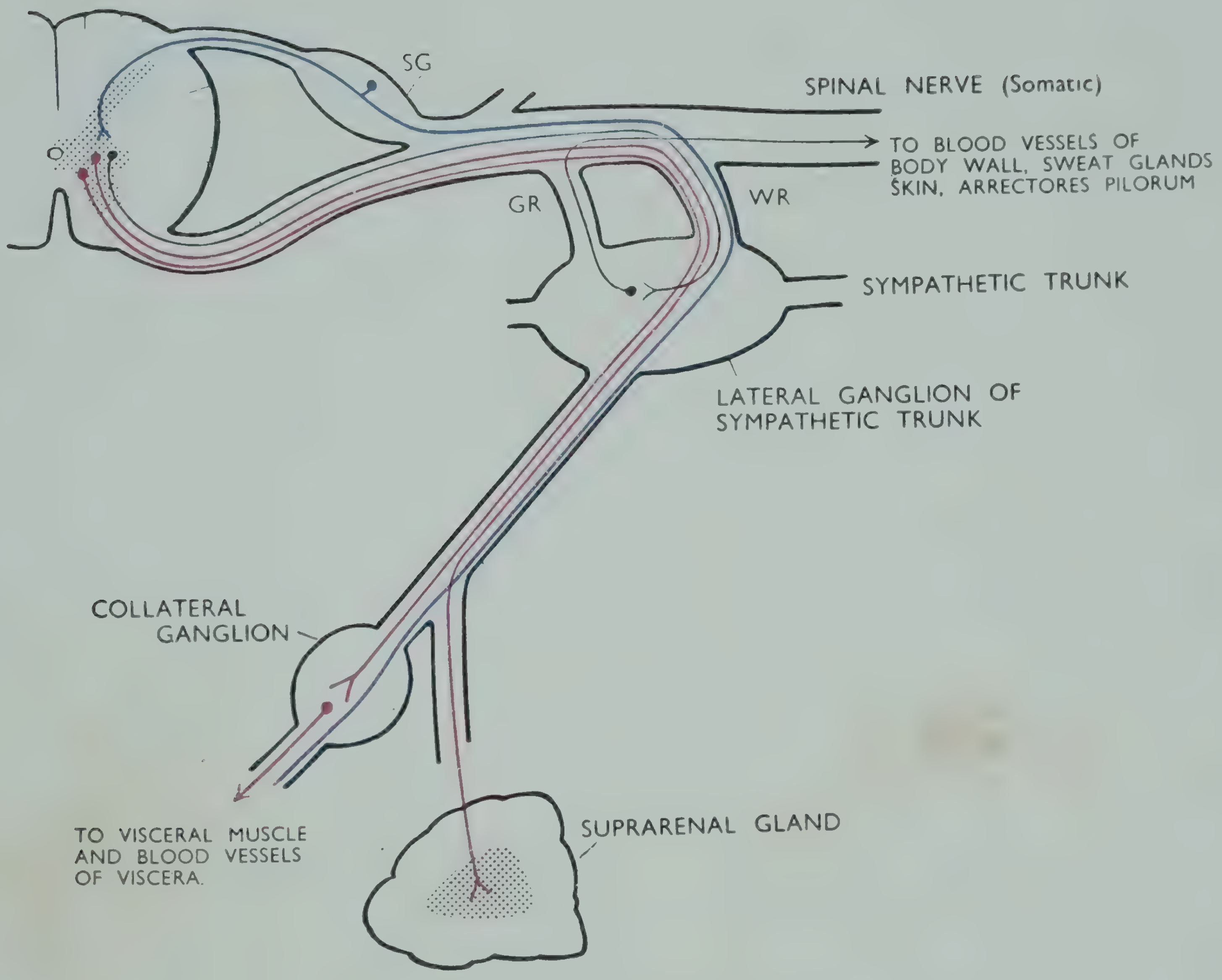

SG
SPINAL NERVE (Somatic)
TO BLOOD VESSELS OF BODY WALL, SWEAT GLANDS SKIN, ARRECTORES PILORUM
GR
WR
SYMPATHETIC TRUNK
LATERAL GANGLION OF SYMPATHETIC TRUNK
COLLATERAL GANGLION
TO VISCERAL MUSCLE AND BLOOD VESSELS OF VISCERA.
SUPRARENAL GLAND

Fig. 43, 2

Pathways of the sympathetic fibres from the lateral horn of the grey matter of the spinal cord. The spinal nerve contains, in addition to postganglionic sympathetic fibres indicated in the diagram, afferent and efferent somatic fibres and is known as a mixed nerve. The connexion between the lateral ganglion and the collateral ganglion could be described as a visceral or splanchnic nerve. The cœliac and hypogastric ganglia are examples of collateral ganglia. The fibre shown in blue represents afferent fibres from the viscera. GR, grey ramus. WR, white ramus. SG, spinal ganglion.

SYMPATHETIC OUTFLOW

The thoracolumbar outflow in man arises from all the thoracic segments of the spinal cord and from the upper two lumbar (Fig. 43, 1); each fibre beginning in a cell of the lateral horn of the grey matter of the spinal cord emerges in an anterior root and leaves the spinal nerve in a *white ramus communicans* which passes to a sympathetic ganglion (Fig. 43, 2). Here the fibre may either end in a synapse or turn aside to run up or down in the sympathetic trunk to a neighbouring ganglion where there is a synapse. From the sympathetic trunk postganglionic fibres, chiefly non-myelinated, pass in the *grey rami communicantes* to every spinal nerve. They are distributed in the peripheral nerves to blood vessels and skin, sweat glands and pilomotor muscles, and in special autonomic nerves (splanchnic nerves) to the abdominal organs. Postganglionic fibres also reach the viscera in the nerve plexus surrounding the arteries which supply them. The suprarenal medulla is exceptional in that it receives preganglionic sympathetic fibres directly from the spinal cord; the cells of the suprarenal gland which produce adrenaline and noradrenaline are homologous with the cells of the peripheral sympathetic ganglia.

The sympathetic system is not essential to life and in human beings large parts of it have been removed without serious impairment of function. Animals, in particular cats, which have undergone as complete a removal as possible of the sympathetic system appear to live normally if they are kept in a protected environment, but they do not survive long if exposed to adverse conditions such as cold, loss of blood or general anæsthesia, the adjustment to which involves complex reactions within the body.

PARASYMPATHETIC (CRANIOSACRAL) OUTFLOW

The cranial part of the parasympathetic outflow is described in other chapters which deal with the organs it supplies. Myelinated preganglionic fibres are found in the third nerve going to the ciliary ganglion, in the seventh going to the submandibular and sphenopalatine ganglia, in the ninth to the otic ganglion, and in the tenth and eleventh cranial nerves to small ganglia and plexuses in the walls of the viscera. The postganglionic fibres are not myelinated and pass to the intra-ocular muscles, salivary and lacrimal glands, heart, lungs and abdominal viscera. The sacral outflow comes from cells situated laterally in the grey matter and leaves the cord by the anterior roots of the second and third sacral nerves as the pelvic splanchnic nerves (nervi erigentes). Their ganglia are usually in the walls of the viscera to which they run. The sacral parasympathetic supplies the colon, rectum, generative organs and urinary bladder. No parasympathetic nerves go to the limbs or to the surface of the body.

AFFERENT IMPULSES FROM THE VISCERA

Running alongside the efferent (outgoing) fibres just described there is an approximately equal number of afferent (ingoing) fibres from the viscera which, like afferents from other parts of the body, have their cell bodies in the spinal ganglia of the posterior roots. Many of these *viscerosensory* fibres form part of visceral reflex arcs but some convey sensation of visceral pain to the brain and all forms of visceral pain can be relieved by sympathectomy. After section of the abdominal viscerosensory nerves the Pacinian corpuscles in the mesentery degenerate and there is apparently no essential difference either in histological or electrical properties between viscerosensory autonomic fibres and afferent fibres from the skin. In the cat impulses in the splanchnic nerves have been traced into the cord, up the posterior columns of white matter on the same side to the nucleus gracilis on the same side and then across the mid-line to the opposite thalamus; more slowly running impulses go up the spinothalamic tracts to the thalamus on both sides and also to the hypothalamus. The function of the pain fibres in man is illustrated by the following observations. If a balloon passed into the small intestine is inflated the subject feels pain in the centre of the abdomen but after bilateral section of the splanchnic nerves, an operation that is sometimes performed for the relief of high blood pressure, inflation of the balloon does not cause pain. Vagotomy has no effect on pain produced by distension of the small intestine. Furthermore, patients with a complete transection of the spinal cord which renders the abdominal wall anæsthetic can not only experience intestinal colic but localize the pain in the centre of the abdomen. Learmonth has reported that a patient who had been given a spinal anæsthetic to anæsthetize the lower part of the body experienced pain accurately localized to the bladder when the superior hypogastric plexus was crushed; the afferent fibres must of course have entered the cord above the level of the anæsthetized part of the cord.

In a conscious, unanæsthetized person the abdominal viscera, provided that they are healthy, can be cut or burned without causing pain; this is clearly demonstrated when a knuckle of colon, brought through the abdominal wall by the surgeon, is opened some days later without any anæsthetic. Wolf and Wolff (p. 233) found that Tom's normal healthy gastric mucosa was not sensitive to pinching or faradic stimulation but if the mucosa became hyperæmic and œdematous pain was readily evoked by these stimuli. In other words, the pain threshold of this viscus is normally high but it may be lowered by inflammatory changes. However, pain arises in the absence of inflammation if the appropriate kind of stimulation is applied. For example, pain is produced if the mesentery is pulled upon or injected with irritant material; since all the nerve fibres from the twenty feet of gut are crowded into the six-inch-long mesentery, traction on this structure must necessarily stimulate a much larger number of nerve fibres than does cutting or pinching of the gut itself. Thus the production of pain by stimulation of visceral nerves depends on (*a*) the strength of the stimulus, (*b*) the number of nerve fibres stimulated and (*c*) the threshold level at the moment.

It is well known that pain arising from a diseased organ may be projected to a definite position but there is no general agreement about the mechanism of the phenomenon. A knowledge of the areas of localization of the pain is, however, of great service in diagnosis and has been summarized in Table 43, 3. The position at which a patient feels pain may coincide with the position of the tissue in which it is produced, or the pain may be felt at a distance from the place of its production. The pain due to pinching the skin on the surface of the body is felt by the subject at the point of stimulation but, when a transposed pedicle skin graft with intact innervation is pinched, the blindfolded patient experiences the

Organ	Superficial Areas to which pain is referred	Segments at which Visceral Afferent Axons enter Spinal Cord																	Peripheral Visceral Pathway
		Thoracic												Lumbar		Sacral			
		1	2	3	4	5	6	7	8	9	10	11	12	1	2	2	3	4	
Heart . . .	Precordium and inner arm	+	+	+	?														Middle and inferior cervical and thoracic cardiac nerves
Lung	No referred pain *																		
Liver and gall-bladder .	Right upper quadrant and right scapula						?	+	+	?									Greater splanchnic nerve
Stomach . . .	Epigastrium						?	+	+	?									Greater splanchnic nerve
Small intestine . .	Umbilicus									+	+	?							Greater splanchnic nerve
Colon { ascending .	Suprapubic											?	+	+					Lumbar chains and pre-aortic plexus
sigmoid and rectum	Deep pelvis and anus															+	+	+	Pelvic nerves and plexuses
Kidney . . .	Loin and groin											?	+	+					Renal plexus via lowest splanchnic nerve and upper lumbar rami
Ureter . . .	Loin and groin													+	+				Renal plexus and upper lumbar rami
Bladder { fundus .	Suprapubic											+	+	+					Hypogastric plexus
bladder neck	Perineum and penis															+	+	+	Pelvic nerves and plexuses
Uterus { fundus . .	Suprapubic region and lower back											+	+	+					Hypogastric plexus
cervix . .	Perineum															+	+	+	Pelvic nerves and plexuses

This table which differs in only minor respects from the one published by Head (1893), has been compiled from all available experimental and clinical data.

---

* Lung parenchyma is insensitive. Pain from larger bronchi is transmitted over somatic vagal axons. When disease spreads to parietal pleura pain is transmitted over intercostal nerves. The afferent axons from the testes enter at thoracic segments 10, 11 and 12; from the prostate and seminal vesicles at T 11 and 12 and L 1.

(From J. C. White, R. H. Smithwick & F. A. Simeone (1952), *The Autonomic Nervous System*, 3rd ed. p. 136. London: Henry Kimpton.)

pain in the position from which the graft was taken and not in its new site. Moreover, pain due to heart disease (angina pectoris) may be projected to the left arm even when this has been amputated. Pain is, therefore, projected to a position and not to any particular tissue. This phenomenon is further illustrated by the fact that pain, produced in organs such as the diaphragm and the testis which have migrated during development, is projected to the primary or original site of the organ. Rubbing the diaphragm at operation or electrical stimulation of the phrenic nerve produces pain which is felt at the root of the neck for the diaphragm develops in the neck, acquires its nerve supply there and later moves caudad. The testis develops near the kidney but, according to Wyndham, low in the abdomen, and therefore testicular pain is felt in the lower abdomen and not in the testis itself, as shown in the following experiment. Scratching of the tunica vaginalis testis gives rise to local pain (as it is comparable with parietal peritoneum) but, after anæsthetization of the tunica, injection of hypertonic saline into the testis produces pain which is localized deeply in the lower abdomen about the level of the internal abdominal ring. In other words the sensorium seems to make contact with, and to become aware of the position of, the organs in embryonic life and it continues to use the same reference map. Gut is a mid-line organ from the embryological point of view and pain arising in it or any other mid-line organ should, according to this interpretation, be projected to various levels of the mid-line no matter whether the part concerned has migrated right or left. This is what is actually found. Thus in colic due to violent peristalsis of the intestine, pain is felt in the mid-line. In disorders of the stomach pain is felt in the mid-line high up in the abdomen; small-intestine pain is felt in the umbilical region; colon pain is situated in the mid-line in the lower part of the abdomen. The gall-bladder is embryologically a mid-line organ and true gall-bladder pain is felt in the epigastrium but, if the inflamed viscus irritates the diaphragm, pain is felt in the neck and right shoulder. It is unlikely that any nerve fibres from the gall-bladder or its peritoneal coverings reach the spinal cord by the phrenic nerve, since the gall-bladder can be painlessly removed under spinal anæsthesia in which the phrenic nerves are functioning normally. Pain arising in the body and cervix of the uterus is felt in the mid-line about one and a half inches above the symphysis pubis.

When irritation of the parietal peritoneum results from visceral disease pain is sharply localized to the site of irritation and is associated with tenderness and often with spasm (rigidity) of the adjacent skeletal muscles. Thus the pain of early appendicitis is felt in the mid-line near the umbilicus and not until the parietal peritoneum, supplied by the somatic nerves, is irritated is the pain felt at the situation of the inflamed appendix, that is, generally, in the right iliac fossa. Similarly, although distension of the gall-bladder produces pain in the centre of the epigastrium, inflammation of the fundus of the gall-bladder is associated with pain in the right hypochondrium in an area overlying the diseased organ.

Visceral pain can be modified by anæsthetization or irritation of the skin area in which the pain is felt. The pain of gastric ulcer is reduced by anæsthetization of the abdominal wall. The pain of angina pectoris can be relieved by anæsthetization of the painful area of the left arm and angina pectoris projected into an amputated limb is relieved by anæsthetization of the brachial plexus. Moreover, if in a patient in whom angina pectoris is elicited by exercise, an area of skin is blistered by cantharidin the pain induced may be felt in the blistered area. Even after complete cutaneous anæsthesia intended to relieve abdominal pain the patient may complain of a dull, aching, deep form of pain which is usually localized in or near to, the diseased organ itself. This pain is similar to that produced by stimulation of the cœliac ganglion and is relieved only by section of sympathetic nerves.

The results of experiments on diaphragmatic pain have been explained by F. R. Brown in the following way. Nerve impulses from the diaphragm and from the shoulder reach the same segment of the spinal cord and possibly the same area of the sensorium (Fig. 43, 4), and by building up the central excitatory state determine the threshold for pain; this may be normally at the level represented at (*a*) in Fig. 43, 4. Anæsthetization of the shoulder area reduces the number of impulses reaching the cord from the shoulder and so raises the threshold as shown at (*b*) and (*c*). Irritation of the diaphragm to a degree represented by the height of the triangle at (*a*) normally produces a shower of impulses which reach the threshold and pain is felt. After anæsthetization of the shoulder, however, a similar degree of irritation of the diaphragm represented by (*b*) now falls short of the threshold and there is no pain. A larger stimulus (large triangle at (*c*)) such as could be produced by a phrenic nerve crush can, however, reach the threshold and cause pain. The observations on angina pectoris already mentioned are also capable of being explained on this theory, since blistering increases the number of afferent impulses arriving at the central nervous system and lowers the threshold for pain.

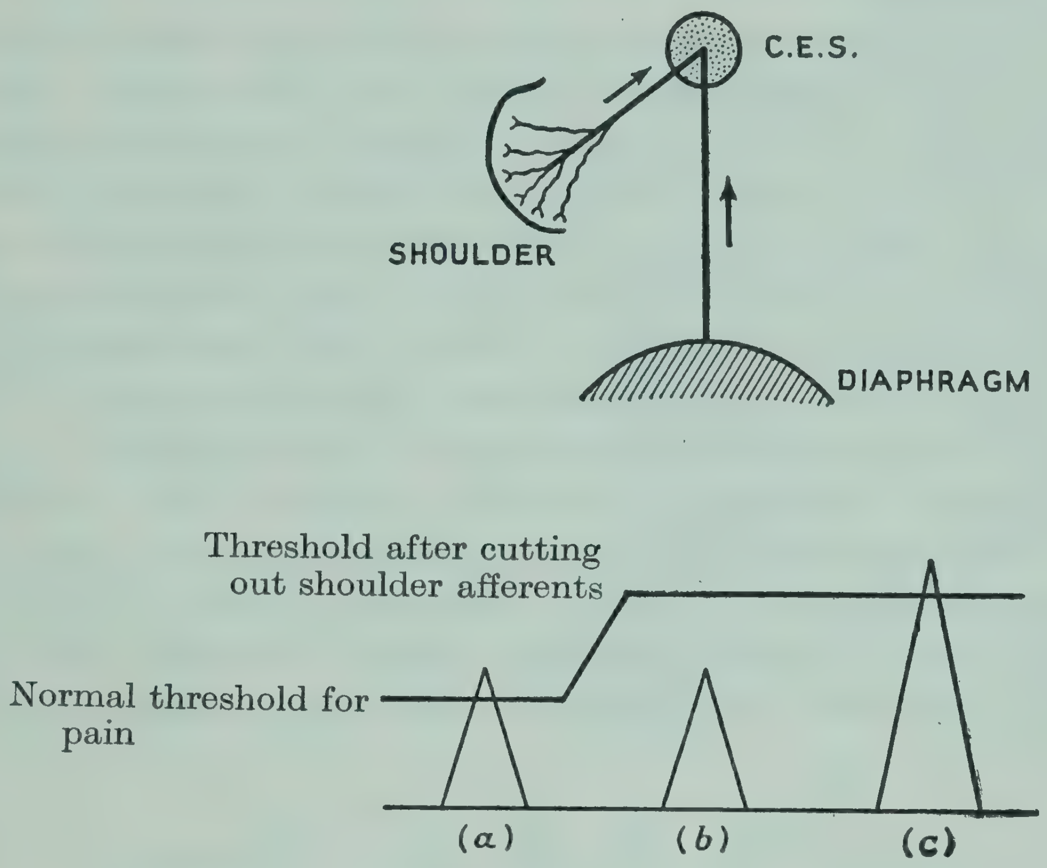

C.E.S.
SHOULDER
DIAPHRAGM
Threshold after cutting out shoulder afferents
Normal threshold for pain
(a)
(b)
(c)

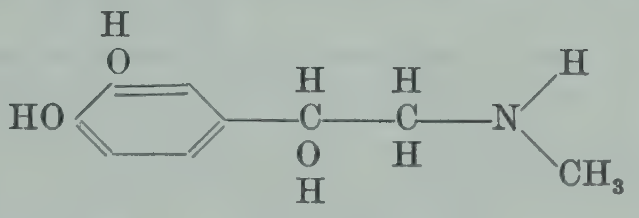
Fig. 43, 4

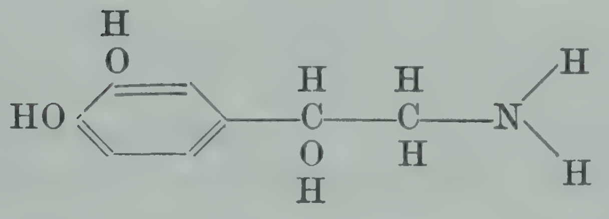
Visceral pain. (*a*) When the diaphragm is rubbed pain is felt in the shoulder. (*b*) When the diaphragm is rubbed after anæsthetization of the shoulder no pain is felt. (*c*) Crushing the phrenic nerve after anæsthetization of the shoulder is painful (F. R. Brown (1948), *Lancet*, i, 386).

HUMORAL THEORY OF TRANSMISSION OF THE EFFECT OF THE NERVE IMPULSE

Since the isolation of adrenaline in 1901 it has been found that there are many similarities between the effects of stimulation of the sympathetic nerves to a part and the effects on it of adrenaline injected into the circulation. This led to the idea, put forward in 1904 by T. R. Elliott, that when an efferent sympathetic nerve is stimulated adrenaline, liberated from some precursor at the nerve-endings, excites the effector organ. Dale in 1934 proposed the name *adrenergic* for those nerves which seem to operate by release of adrenaline.

Certain discrepancies between the effects of adrenaline and of sympathetic stimulation have been known for some time. For example in the cat stimulation of the hepatic or splenic nerves produces an increase in the rate of the denervated heart but does not relax the uterus, while the injection of adrenaline accelerates the heart and relaxes the uterus. These discrepancies can now be accounted for. The substance L-noradrenaline, although synthesized as early as 1904, did not become available for experimental purposes until 1948 when pharmacological methods for its identification in blood and tissues began to be developed. It is now clear that sympathetic adrenergic nerves contain both noradrenaline and adrenaline, with noradrenaline preponderating, and that on stimulation they release simultaneously both adrenaline and noradrenaline. In all cases the amount of noradrenaline released exceeds that of adrenaline. Noradrenaline as well as adrenaline has been found in normal human suprarenal glands and the material discharged from the suprarenal medulla in response to stimulation of the splanchnic nerve in most mammals is a mixture of adrenaline and noradrenaline. It is likely that noradrenaline is a precursor of adrenaline, into which it is converted by the process of methylation (p. 347). Noradrenaline is a powerful vasoconstrictor. If it is given by intravenous infusion in man, in doses of the order of 20 μg. per min, both the systolic and the diastolic blood pressures are raised and in consequence there is produced reflexly cardiac slowing so that the cardiac output is little changed (Fig. 30, 10). Adrenaline, on the other hand, can in small doses augment the cardiac output without raising the blood pressure (Fig. 28, 37). If larger doses of adrenaline of the order of 20 μg. per min are infused, acceleration of the heart with increase of systolic pressure but no increase of diastolic blood pressure is observed. The explanation of these differences is that adrenaline causes a dilatation of the vessels in the skeletal muscles and therefore a reduction of peripheral resistance. Noradrenaline by constricting the vessels in the muscles as


adrenaline


noradrenaline

well as those in the skin and viscera causes a rise in the total peripheral resistance. The actions of the two substances are compared in Table 43, 5, which shows that on the whole the two substances have the same effect, perhaps to different degrees. The word 'sympathin' is used by some authors as a convenient description of the mixture of adrenaline and noradrenaline which is released at sympathetic nerve endings. The adrenaline-like compounds, produced on stimulation of a sympathetic nerve, enter the circulation since they are only relatively slowly destroyed in the tissues by amine oxidase and catechol oxidases. Their presence in the blood is shown by acceleration of the heart or withdrawal of the nictitating membrane, provided that these structures have been denervated about a week previously. Denervation makes tissues which are supplied by sympathetic nerves more sensitive to adrenaline and even more sensitive to noradrenaline. It is possible that the increase in sensitivity depends on the reduction in the amount of amine oxidase in the endings subsequent to degeneration of the nerve. Since this enzyme acts more

TABLE, 43, 5

Table to compare the Effects of L-*Adrenaline and* L-*Noradrenaline*

	Effect of	
	L-adrenaline	L-noradrenaline
Man		
Heart rate	Increase	Decrease
Cardiac output	Increase	Variable
Total peripheral resistance	Decrease	Increase
Blood pressure	Rise	Greater rise
Respiration	Stimulates	Stimulates
Skin vessels	Constriction	Less constriction
Muscle vessels	Dilatation	Constriction
Bronchus	Dilatation	Less dilatation
Eosinophil count	Increase	No effect
Metabolism	Increase	Slight increase
Oxygen consumption	Increase	No effect
Blood sugar	Increase	Slight increase
Central nervous system	Anxiety	No effect
Uterus *in vivo* in late pregnancy	Inhibits	Stimulates
Animals		
Blood pressure (cat and dog)	Rise	Greater rise
Spleen perfused *in situ* (cat)	Contraction	Contraction
Uterus pregnant (cat and rabbit)	Contraction	Contraction
Uterus nonpregnant (rabbit)	Contraction	Contraction
Uterus nonpregnant (cat)	Inhibition	Less inhibition
Ileum isolated (rabbit)	Inhibition	Less inhibition
Bronchus (guinea-pig and dog)	Dilatation	Less dilatation
Eye—pupil	Dilatation	Less dilatation
nictitating membrane (cat)	Contraction	Contraction
Kidney	Vasoconstriction	Vasoconstriction
Isolated heart—rate and force	Increase	Increase
Coronary vessels	Dilatation	Dilatation

rapidly on noradrenaline than on adrenaline the greater potentiation of the effect of noradrenaline after denervation is thus explained. Ephedrine, by inhibiting the action of the amine oxidase, potentiates the action of adrenaline and noradrenaline and also increases the effects obtained by stimulation of adrenergic nerves. In animals, administration of thyroid hormone reduces the amine oxidase content of the liver and blood vessels and this may account for the exaggerated hyperglycæmia and rise of blood pressure after adrenaline injection in hyperthyroidism.

Organs supplied by postganglionic adrenergic nerves contain noradrenaline provided their nerve supply is intact; after denervation the noradrenaline (but not adrenaline) almost entirely disappears but returns when the nerves regenerate. The nerve fibres must, therefore, be the only important source of noradrenaline. It is found that prolonged stimulation of an adrenergic nerve releases nonadrenaline at a steady rate and does not reduce the noradrenaline content of the organ supplied; it must be assumed, therefore, that resynthesis is rapid enough to keep pace with the loss. The wall of blood vessels is rich in noradrenaline and the relative abundance of nerve fibres can be judged by the quantity of noradrenaline they contain. The greater the supply of nerve fibres the more vigorously do they contract under the influence of noradrenaline. In conformity with this generalization the vessels of the skin and spleen, for example, react strongly to noradrenaline while those of the lung and brain respond only feebly.

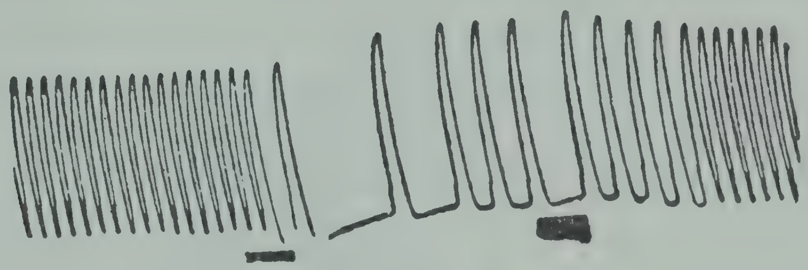
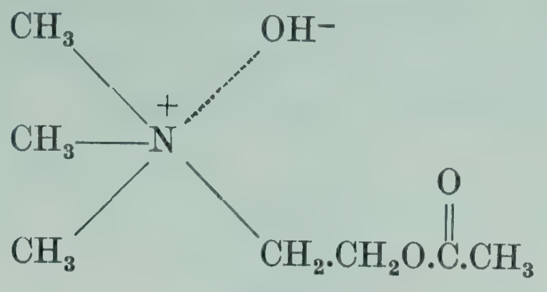


Fig. 43, 6

Unpublished record from an experiment in 1906 by the late W. E. Dixon. Beat of the exposed heart of a frog. At the first mark extract from inhibited dog's heart applied; at the second mark atropine (H. H. Dale (1934), *Brit. med. J.* i, 835).

Evidence has accumulated in a somewhat similar fashion about the substance released when parasympathetic nerves are stimulated. In 1906 Dixon found that after vagal stimulation an extract of heart tissue caused an inhibition of a frog's heart; the inhibition could be prevented by atropine (Fig. 43, 6). Interest was directed to these results when Loewi in 1921 showed that, during stimulation of the vagus, Ringer's solution perfusing a frog's heart acquired a substance (vagus substance) which could inhibit another frog's heart. Later work showed that the substance released from the heart is acetylcholine and nerves acting in this way are described by Dale as *cholinergic*.

Stimulation of certain sympathetic fibres, for example sympathetic vasodilator nerves to the muscles of the dog and also the cat, liberates acetylcholine. These nerves, then, come into Dale's category of cholinergic nerves. Stimulation of the stellate ganglion causes an acetylcholine-like substance to appear in the coronary perfusate; presumably the sympathetic vasodilator fibres in the coronary arteries liberate acetylcholine.



acetylcholine

The sympathetic nerves to the suprarenal medulla liberate acetylcholine which in turn calls forth a secretion of adrenaline and noradrenaline. This is not however to be regarded as an exception to the rule that the sympathetic nerves are adrenergic, since the suprarenal medulla is derived in the embryo from a tissue which forms the sympathetic ganglia and the medulla is therefore homologous with a sympathetic ganglion in which the transmitter is always acetylcholine. Although the fibres to the sweat glands are in fact postganglionic sympathetic fibres they act in most animals, including man, by releasing acetylcholine, that is they are cholinergic.

Cholinergic nerves were first demonstrated in the frog. In mammals, however, it was not possible to demonstrate that certain nerves acted by producing acetylcholine at their endings until it was discovered that both blood and tissues contain cholinesterase, the action of which is blocked by eserine (physostigmine). After its release from the nerve-endings acetylcholine is quickly hydrolysed by the esterase to choline and acetic acid and its action is, therefore, local and transient. Only after an animal is eserinized does the acetylcholine released on parasympathetic nerve stimulation enter the circulation in sufficient quantity to produce effects at a distance. For example, stimulation of the chorda tympani nerve normally produces a vasodilatation confined to the submandibular and the sublingual glands and a secretion of saliva without any alteration in the general blood pressure, but after an intravenous injection of eserine these effects are accompanied by a lowering of blood pressure due to a general vasodilatation. Extremely sensitive and fairly specific pharmacological tests have been found to detect minute amounts of acetylcholine in the fluid or blood which has perfused an organ. The eserinized muscle of the leech, for example, can detect 1 part in 10^9 of acetylcholine. The site of production of acetylcholine seems to be the nerve-endings and not the tissue cells, since after denervation the acetylcholine content of an organ declines as the nerves degenerate and finally disappears. Acetylcholine is stored in the nerve in an inactive form, probably bound to protein, which is not affected by cholinesterase, and it is only after its release from the nerve-ending as free acetylcholine that hydrolysis by cholinesterase occurs. A cholinergic neurone contains acetylcholine along the whole course of the axon as well as at its endings. Acetylcholine acts on the effector tissue by depolarizing its cells. Denervation of tissues supplied by cholinergic nerves sensitizes them to acetylcholine (Fig. 42, 20).

We are accustomed to think of events in the nervous system in terms of milliseconds and it is difficult at first to fit in this rapidity of action with humoral transmission in which acetylcholine must diffuse from, say, the preganglionic nerve endings to the ganglion cell. The distance between endings and ganglion cell is, however, so small that there is no difficulty in explaining the synaptic delay of a few milliseconds. Furthermore, acetylcholine is so quickly destroyed by cholinesterase that the acetylcholine released by one nerve impulse is reduced to ineffective amounts before the next impulse arrives at the synapse.

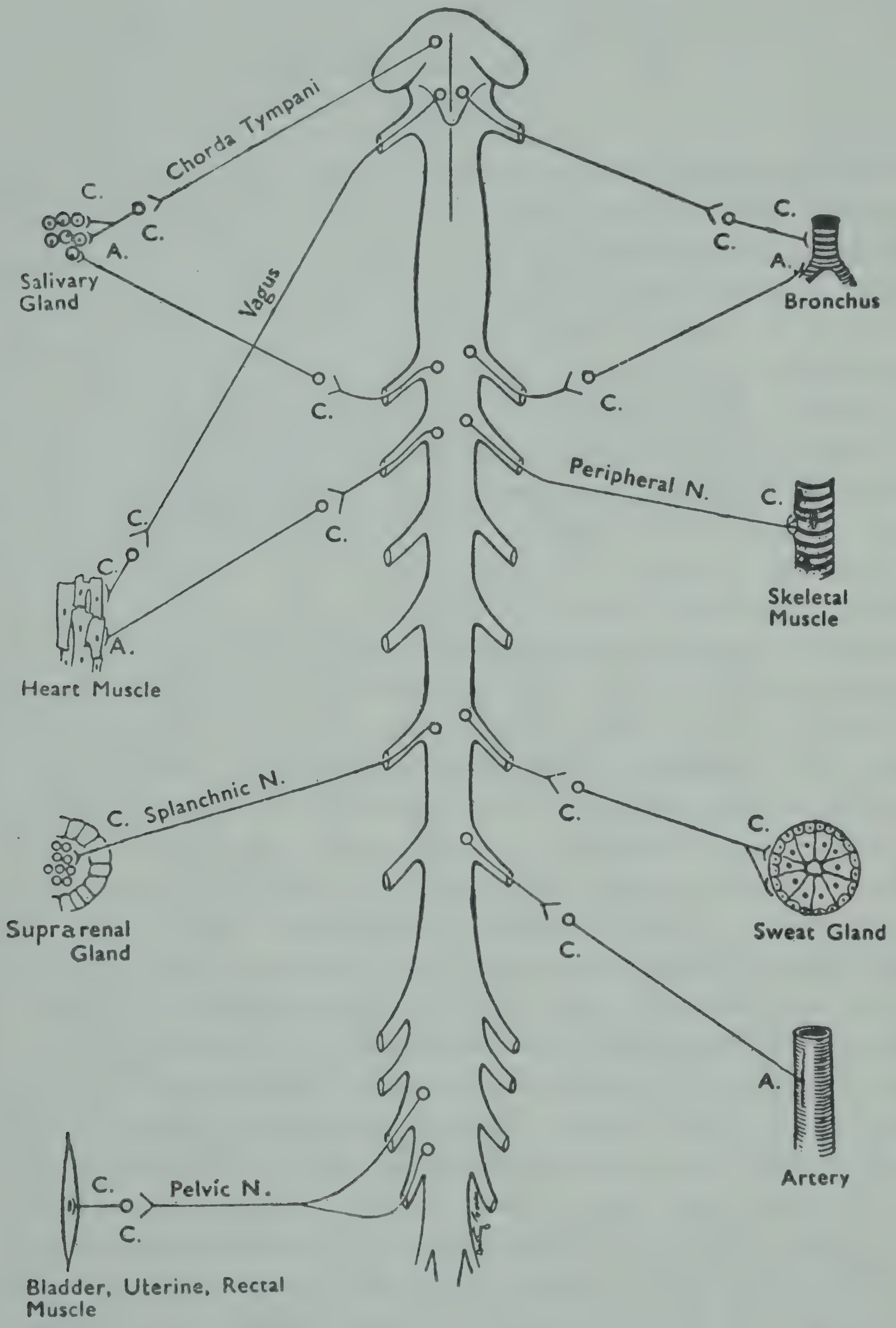

Chorda Tympani
C.
C.
A.
Salivary Gland
Vagus
C.
C.
C.
A.
Bronchus
C.
C.
C.
Peripheral N.
C.
Skeletal Muscle
C.
C.
C.
A.
Heart Muscle
C. Splanchnic N.
Suprarenal Gland
C.
C.
Sweat Gland
C.
A.
Artery
C.
Pelvic N.
C.
Bladder, Uterine, Rectal Muscle

Fig. 43, 7

Adrenergic and cholinergic nerve-endings. A, adrenergic nerve-ending. C, cholinergic nerve-ending. Redrawn and modified from Dale (1934), *Brit. med. J.* i, 835. (From J. C. White, R. H. Smithwick & F. A. Simeone (1952), *The Autonomic Nervous System*, 3rd ed. p. 96. London: Henry Kimpton.)

A diagram of the distribution of adrenergic and cholinergic nerves is given in Fig. 43, 7. Where a tissue receives both sympathetic and parasympathetic fibres it is often found that the nerve impulses arriving by one system have the opposite effects to nerve impulses arriving by the other. For example, vagal impulses (parasympathetic) cause slowing of the heart and sympathetic impulses cause acceleration. Since such inhibitory and excitatory differences are almost universally found in the body it must be assumed that there is normally some kind of balance between the two systems. In other words, the autonomic system is a homoiostatic mechanism (Fig. 43, 8). If the sympathetic system is eliminated by removal of the sympathetic trunks this balance is upset and the animal survives only if it is sheltered from situations which put too great a strain on its means of physiological adjustment.

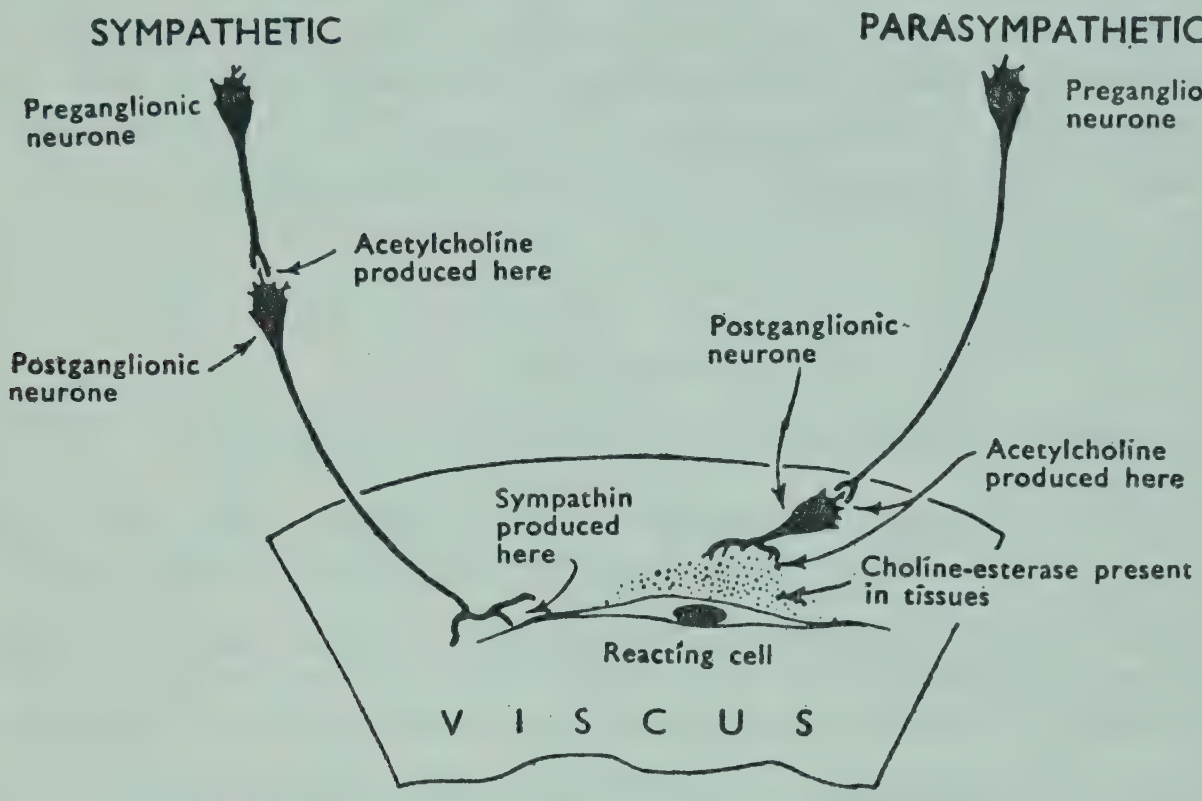

SYMPATHETIC
PARASYMPATHETIC
Preganglionic neurone
Preganglionic neurone
Acetylcholine produced here
Postganglionic neurone
Postganglionic neurone
Acetylcholine produced here
Sympathin produced here
Choline-esterase present in tissues
Reacting cell
VISCUS

FIG. 43, 8

Balance between sympathetic and parasympathetic nerves and esterase. Sympathin is a term used for the material released at sympathetic nerve-endings (A. Myerson (1938), *J. Amer. med. Assoc.* **110,** 101).

The importance of the autonomic nervous system for homoiostasis in man is illustrated by observations, especially those of Guttmann and Whitteridge, on men who survive in good health after transverse lesions of the spinal cord. Consideration of Fig. 43, 1, shows that a lesion above the first thoracic segment (T1) cuts off the whole of the thoracolumbar autonomic outflow from higher control. Tilting such a patient suddenly into the upright position causes a fall of blood pressure, a rise in pulse rate and loss of consciousness. The absence of the normal compensatory vasoconstriction is made obvious by the legs becoming purple. The patient learns to get into the upright position slowly—perhaps the cells of the brain can become adapted to a slow fall of blood pressure. Such a patient, since he has lost control of skin vessels, can produce neither vasoconstriction nor vasodilatation nor sweating to keep the body temperature steady; if he becomes very cold he cannot shiver as the muscles below the lesion are paralysed. These disabilities obviously become smaller the lower the lesion in the spinal cord. When the urinary bladder of a paraplegic patient is distended there is constriction of the vessels of the skin of the legs and of the muscles; in a high transverse lesion the systolic blood pressure may rise as high as 260 mm Hg and the diastolic as high as 140 mm Hg. In a warm environment the body temperature may rise steadily. The rise of blood pressure is accompanied by flushing of the face with sweating and headache. In low lesions the flushing is presumably due to reduction of outflow of vasoconstrictor impulses in sympathetic nerves still in connexion with higher centres; but flushing of the face occurs even with lesions above the sympathetic outflow, perhaps as an axon reflex. In lesions below T6 bladder distension produces as marked a vasoconstriction of the legs but in spite of this there is relatively little effect on the blood pressure. This shows that with lower lesions the compensatory mechanisms are effective. The vagal reflexes which cause slowing of the heart, when the blood pressure rises, are intact and this may in part account for the maintenance of homoiostasis in low transverse lesions of the cord. Effects similar to those just described for bladder distension may be produced by distension of the rectum.

REFERENCES

Brown, G. L. (1937). Transmission at nerve-endings by acetylcholine. *Physiol. Rev.* **17,** 485–513.

Burn, J. H. (1952). The enzyme at sympathetic nerve-endings. *Brit. med. J.* i, 784–787.

von Euler, U. S. (1951). Hormones of the sympathetic nervous system and the adrenal medulla. *Brit. med. J.* i, 105–108.

von Euler, U. S. (1955). Noradrenaline in hypotensive states and shock: physiological aspects. *Lancet,* ii, 151–154.

Feldberg, W. (1952). Physiology of the autonomic nervous system. *Lect. Sci. Bas. Med.* **2,** 343–374.

Kuntz, A. (1949). *The Neuroanatomic Basis of the Surgery of the Autonomic Nervous System.* Oxford: Blackwell.

Lands, A. M. (1949). The pharmacological activity of epinephrine and related dihydroxyphenylalkylamines. *Pharmacol. Rev.* **1,** 279–309.

Learmonth, J. (1950). The surgery of the sympathetic nervous system. *Lancet,* ii, 505–508.

Metabolism and function in nervous tissue (1952). *Biochemical Soc. Symp. No. 8.*

Mitchell, G. A. G. (1953). *Anatomy of the Autonomic Nervous System.* Edinburgh: Livingstone.

Nickerson, M. (1949). The pharmacology of adrenergic blockade. *Pharmacol. Rev.* **1,** 27–101.

Paton, W. D. M. (1951). Paralysis of autonomic ganglia and the therapeutic effects of ganglion blocking drugs. *Brit. med. J.* i, 773–778.

Paton, W. D. M. (1953). Anticholinesterases. *Lect. Sci. Bas. Med.* **3,** 105–126.

Rosenblueth, A. (1950). *The Transmission of Nerve Impulses at Neuroeffector Junctions and Peripheral Synapses.* London: Chapman & Hall.

White, J. C., Smithwick, R. H. & Simeone, F. A. (1952). *The Autonomic Nervous System.* 3rd ed. London: Henry Kimpton.

Whittaker, V. P. (1951). Specificity mode of action and distribution of cholinesterase. *Physiol. Rev.* **31,** 312–343.

Whitteridge, D. (1954). Cardiovascular disturbances in paraplegics. *Edinb. med. J.* **61**, 1–6.

CHAPTER 44

NERVOUS SYSTEM

INTRODUCTION

IT is not intended in this book to give a detailed account of the anatomy of the nervous system but only to supply sufficient information to make the main pathways and connexions understandable. We shall begin with a general account to introduce the names of the chief parts.

The nervous system can be divided into *central* and *peripheral* portions. The central nervous system consists of the brain contained within the skull and the spinal cord lying in the vertebral canal. The peripheral part consists of forty-three pairs of nerves which leave the cerebrospinal axis and pass to the various organs of the body. The peripheral nerves contain fibres of two kinds, one group termed *afferent* or *sensory* fibres carries ingoing nerve impulses from the periphery to the central nervous system. The second group, the *efferent* or *motor* fibres, carries outgoing impulses from the central nervous system to muscles and glands and other organs. Each spinal nerve arises from the cord by two roots; the posterior (dorsal in animals) carries afferent fibres while the anterior (or ventral) is occupied by efferent fibres.

The nervous system can also be divided into two anatomical and functional components, the *somatic* (or cerebrospinal) and the *autonomic* (or visceral). The somatic component has a peripheral part, the nerves of the body, and a central part, the brain and spinal cord, and is concerned both with information arriving from the surface of the body, from the muscles and joints, and with the transmission of impulses to the muscles. The autonomic component may also be divided into a central part consisting of nerve cells and their processes which lie in the central nervous system and a peripheral part made up of fibres passing to and from the blood vessels and viscera.

The nervous system may be regarded as the main co-ordinator of the activities of the body. The afferent nerves bring to the central nervous system information about the external world from receptors which are sensitive to stimuli such as light, sound, temperature, or pressure, and also information about the internal state of the body (for example tension in muscles, distension in viscera). As a consequence of the arrival of these impulses the central nervous system sends impulses along the efferent nerves to produce appropriate movements of muscles or secretion of glands. The pathways taken by impulses after they arrive at the central nervous system are very complex, but the basic principle is that impulses are handed on from nerve cell to nerve cell until finally they emerge in an efferent nerve to produce a response in an effector organ such as a muscle or a gland. Sometimes this process rises to conscious levels but many of the activities of the body are regulated without our being aware of them. These activities controlled by the nervous system are called *reflex actions*, or simply *reflexes*, and the pathways followed in the nervous system along afferent, junctional and efferent neurones are called *reflex arcs*.

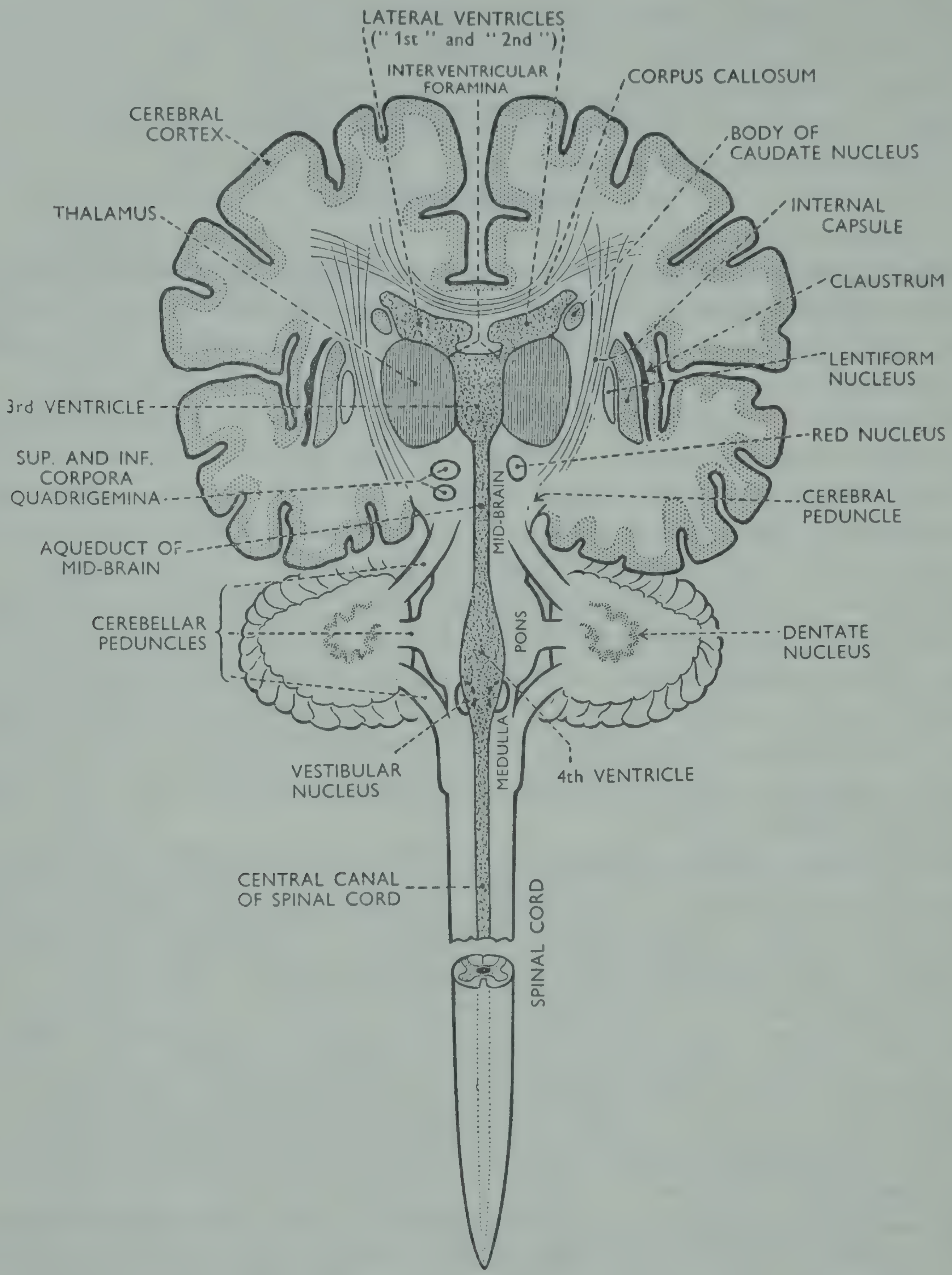
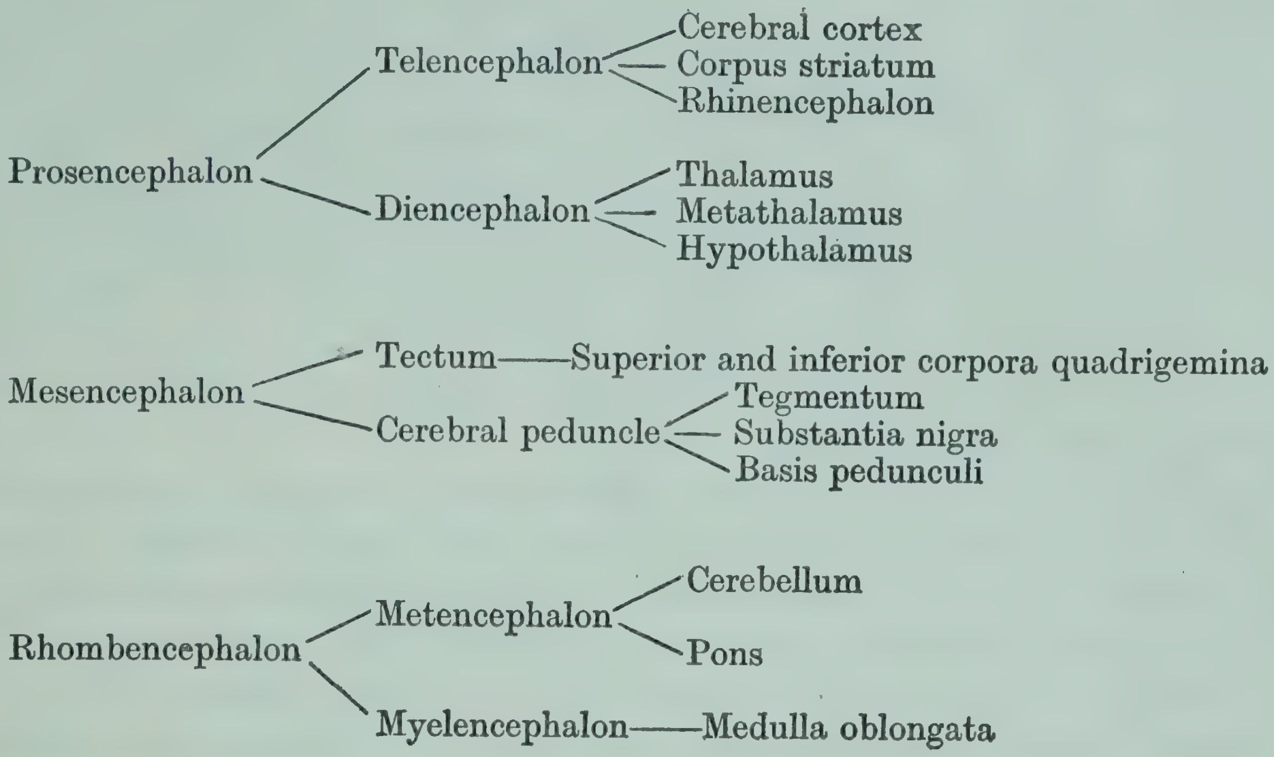

LATERAL VENTRICLES
("1st" and "2nd")
INTERVENTRICULAR FORAMINA
CORPUS CALLOSUM
CEREBRAL CORTEX
BODY OF CAUDATE NUCLEUS
THALAMUS
INTERNAL CAPSULE
CLAUSTRUM
LENTIFORM NUCLEUS
3rd VENTRICLE
RED NUCLEUS
SUP. AND INF. CORPORA QUADRIGEMINA
CEREBRAL PEDUNCLE
MID-BRAIN
AQUEDUCT OF MID-BRAIN
CEREBELLAR PEDUNCLES
PONS
DENTATE NUCLEUS
MEDULLA
VESTIBULAR NUCLEUS
4th VENTRICLE
CENTRAL CANAL OF SPINAL CORD
SPINAL CORD

FIG. 44, 1

Diagrammatic coronal section through the cerebral hemispheres, brain-stem and spinal cord to show the approximate relationships of the more important parts. Some of the structures are shown on one side only and the two halves of the cerebellum have been separated as they were in early development. The caudate and lentiform nuclei together form the corpus striatum.
(*By courtesy of J. D. B. MacDougall.*)

The cerebrospinal axis is developed out of a simple tube of ectoderm. The lumen of the cranial end of this neural tube becomes dilated and eventually produces the large ventricular system within the brain; the lumen of the caudal part of the tube remains very small and is recognizable in the adult as the tiny central canal of the spinal cord. In the human embryo about five weeks old three main dilatations of the cranial or cephalic end of the neural tube can be made out: 1, prosencephalon or forebrain vesicle; 2, mesencephalon or midbrain vesicle; and 3, rhombencephalon or hindbrain vesicle. The more important parts of the brain formed from these vesicles are shown in the following schema, and a diagram of the main parts of the brain is given in Fig. 44, 1.



- Prosencephalon
  - Telencephalon
    - Cerebral cortex
    - Corpus striatum
    - Rhinencephalon
  - Diencephalon
    - Thalamus
    - Metathalamus
    - Hypothalamus
- Mesencephalon
  - Tectum——Superior and inferior corpora quadrigemina
  - Cerebral peduncle
    - Tegmentum
    - Substantia nigra
    - Basis pedunculi
- Rhombencephalon
  - Metencephalon
    - Cerebellum
    - Pons
  - Myelencephalon——Medulla oblongata

In the descriptions of the anatomy of the nervous system which follow it is necessary to use certain technical terms. When referring to the spinal cord in the human subject the terms *anterior* and *posterior* have the usual anatomical meanings; in the case of animals *ventral* and *dorsal* are used. The *median plane* passes anteroposteriorly in man and divides the body into two symmetrical halves; any plane parallel to the median plane is called a *sagittal plane*. A vertical plane at right angles to this, for example a vertical plane passing through both external auditory meati, is called a *coronal plane*. It is convenient to refer to points in one vertical coronal plane as being above or dorsal to points in the same plane lying below or ventral. In the central nervous system the cord can be regarded as lower than the medulla, which is lower than the mid-brain, and the cerebral hemispheres are then referred to as the highest parts. To avoid any implication which may be contained in the words higher or lower the mid-brain may be described as being *rostral* or *cephalic* to the medulla which is the most *caudal* part of the brain. The adverbial forms *cephalad* (towards the rostral or cephalic end of the brain) and *caudad* (towards the tail end of the brain) are in common use.

The brains of primitive animals are much simpler than those of animals higher up the evolutionary scale but quite often certain parts of the primitive brain, a collection of nerve cells for example, can be recognized in higher animals because of a similarity in position or in relationship to other structures. Such areas are designated *palæo*. Later developments in the evolutionary process with no counterparts in the primitive brain are designated with the prefix *neo*.

METABOLISM OF THE NERVOUS SYSTEM

The brain receives its food materials almost entirely from the blood stream, very little exchange occurring by way of the cerebrospinal fluid. The total blood flow through the brain in man has been estimated by a method involving the inhalation of nitrous oxide and is on the average 65 ml. per 100 g. per minute. If the average brain weight is taken as 1400 g. the flow is about 900 ml. per min. The blood leaving the brain contains rather less oxygen than the mixed venous blood and the carotid-venous oxygen difference is about 6 ml. per 100 ml. of blood. The oxygen consumption of the brain is thus $\frac{65}{100} \times 6$ or nearly 4 ml. oxygen per 100 g. brain tissue or 56 ml. of oxygen per minute for the whole brain. This is about one fifth of the basal oxygen requirements for the whole body. In infants with their relatively larger brains it is to be expected that the proportion of the total oxygen consumption needed by the brain is much higher, perhaps as much as 50 per cent.

By comparing the arterio-venous differences in oxygen, carbon dioxide and glucose it has been found that the R.Q. of brain is about 1 and that the oxygen utilized is equivalent to the amount of glucose which disappears. Brain tissue appears therefore to make extensive use of glucose as a source of energy (the glycogen content is probably too low, about 0·1 per cent., to make an important contribution), and this conclusion is confirmed by the observation that in hypoglycæmia due to insulin overdosage the oxygen consumption shows a sharp fall.

A large part of the glucose utilized by the brain is broken down by glycolysis (Fig. 20, 17) to pyruvic and lactic acids. Of these an amount corresponding to about 15 per cent. of the glucose consumed appears in the venous blood leaving the brain; the remainder not removed by the blood is oxidized to carbon dioxide and water by the tricarboxylic acid cycle (Fig. 20, 18). For the entry of pyruvic acid into this cycle the participation of thiaminediphosphate (cocarboxylase) is essential. It is not surprising therefore to find that, as Peters first showed, in thiamine deficiency pyruvic acid accumulates in brain tissue. Some of the glucose utilized in brain is probably also metabolized by the direct oxidative pathway mentioned on page 316.

While therefore brain tissue possesses all the mechanisms for oxidative phosphorylation and while it undoubtedly uses glucose as a source of energy in the production of ATP, it is important to note that none of the intermediates in the glycolytic or tricarboxylic acid cyles is effective in combatting hypoglycæmia. For this purpose glucose seems to have a special and specific role.

Since it obtains so much of its energy requirements by the oxidation of glucose either directly or through the tricarboxylic acid cycle and since it has such a high oxygen consumption, it will be readily understood that the action of the brain is easily disturbed by anoxia, even for very short periods.

The only amino-acid capable of being metabolized by brain is glutamic acid which is converted to α-ketoglutaric acid which in turn enters the tricarboxylic acid. Glutamine is also utilized, being converted to glutamic acid by the enzyme glutaminase.

REFERENCES

FULTON, J. F. (1949). *The Physiology of the Nervous System.* Oxford University Press.

McILWAIN, H. (1955). Biochemistry and the central nervous system. London: Churchill.

Metabolism and function in nervous tissue (1952). *Biochem. Soc. Symp. No. 8.* Cambridge University Press.

SCHMIDT, C. F. (1950). *The Cerebral Circulation in Health and Disease.* Oxford: Blackwell.

SHERRINGTON, C. S. (1947). *The Integrative Action of the Nervous System.* Cambridge University Press.

WEIL-MALHERBE, H. (1950). Significance of glutamic acid for the metabolism of nervous tissue. *Physiol. Rev.* **30,** 549–568.

WEISS, P. (Editor) (1950). *Conference on "Genetic Neurology." Problems of the Development, Growth and Regeneration of the Nervous System and of its Functions.* University of Chicago Press.

CHAPTER 45

SPINAL CORD

THE PATHWAYS OF THE SPINAL CORD

THE spinal cord (medulla spinalis) extends in adult man from the upper border of the first cervical vertebra (the atlas) to the lower border of the first lumbar vertebra. It is about 1 cm in diameter on the average, being a little wider at the cervical and lumbar enlargements where the nerves to the arms and legs arise, because of the greater amount of grey matter there. There is no trace of segmentation on the surface of the spinal cord but it is convenient to describe it as if it were divided up into thirty-one segments from each of which arises a pair of *spinal nerves* (Fig. 45, 1).

The *posterior root* (dorsal root in animals) emerges from the postero-lateral aspect of the segment as a number of rootlets which fuse as they pass laterally to form a single trunk upon which is an enlargement, the *spinal* or *posterior root ganglion.* The *anterior root* (ventral root in animals) is formed by the fusion of a number of rootlets which arise in an irregular fashion from the antero-lateral aspect of the cord. The anterior root joins the posterior root beyond the ganglion to form the spinal nerve which emerges between the vertebræ.

The interior of the cord around the central canal is occupied by an H-shaped mass of *grey matter* consisting largely of nerve cells (Fig. 45, 1). The shape of the grey matter varies from segment to segment but it shows three main parts, anterior, lateral and posterior columns or horns. From the large motor cells of the anterior horn axons pass in the anterior or motor root to the voluntary muscles. The small lateral horns, seen only in the thoracic and upper lumbar regions of the cord, contain cells of medium size, belonging to the sympathetic system, from which axons travel in the anterior roots to the white rami communicantes (p. 805). The posterior column of grey matter shows two main collections of small nerve cells, the thoracic nucleus at the base and the substantia gelatinosa at the apex of the horn. Scattered throughout the grey matter of the posterior horn are other cells which form the nucleus proprius and give off fibres to connect neighbouring segments.

The peripheral part of the cord, the *white matter*, consists almost entirely of myelinated nerve fibres supported in a meshwork of neuroglia. For descriptive purposes the white matter can be divided into posterior, lateral and anterior columns as shown in Fig. 45, 1, although these areas are not very precisely marked off from one another. In the white matter nerve fibres having similar functions and destinations are grouped together but since the various groups cannot be distinguished by their microscopic appearance special methods are required to identify them.

METHODS OF DETERMINING ANATOMICAL PATHWAYS

The degeneration which is a consequence of the cutting of a peripheral nerve has already been described (p. 763). The mapping out of the degeneration after section or destruction of a part of the nervous system is a most important method for finding the routes taken by nerve fibres in the central nervous system. Experiments of this kind have been made very frequently in animals but, in view of the differences between the nervous system of man and that of the lower animals, degenerations produced in man by surgical operation or by war wounds give more valuable information.

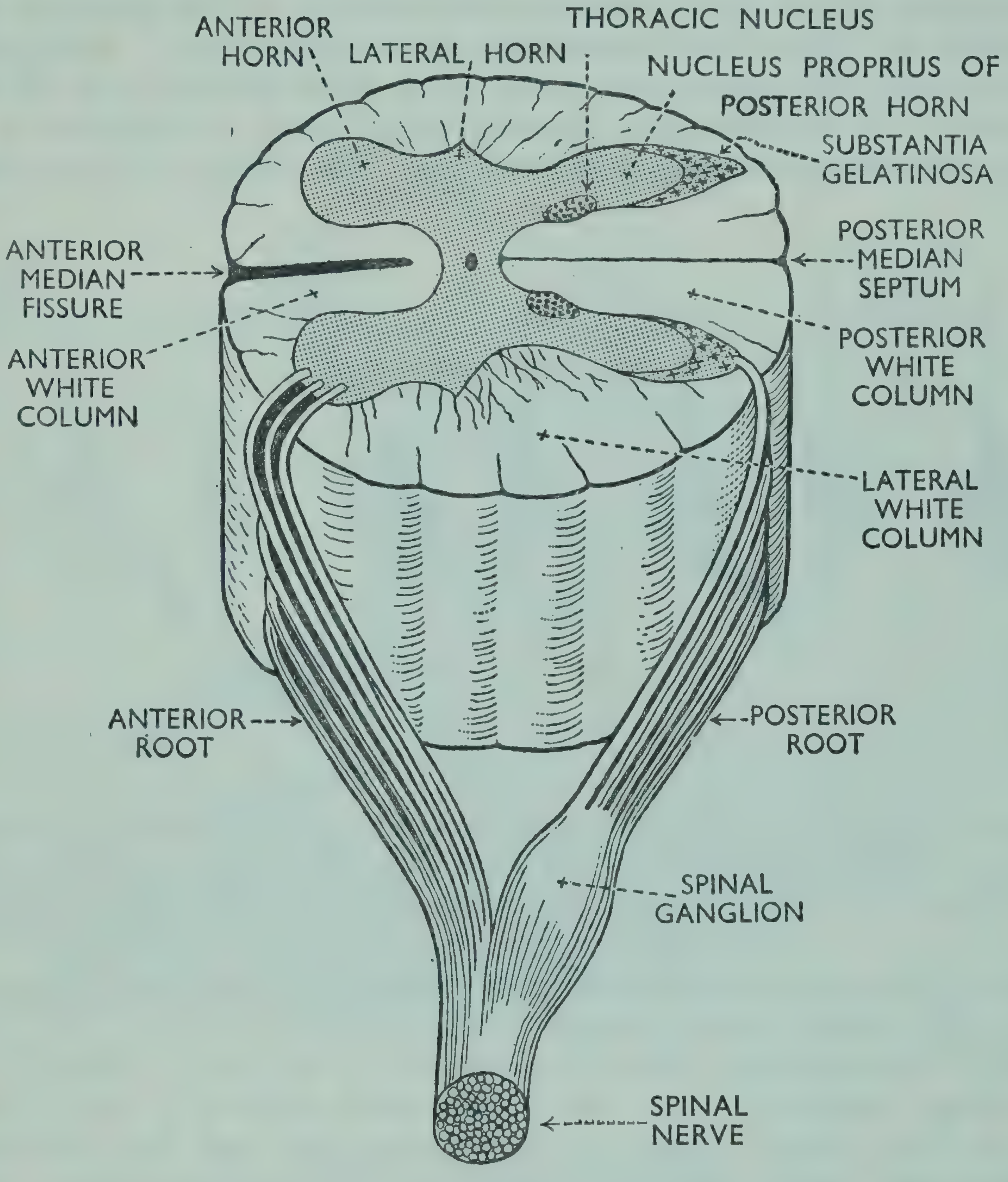

ANTERIOR HORN
LATERAL HORN
THORACIC NUCLEUS
NUCLEUS PROPRIUS OF POSTERIOR HORN
SUBSTANTIA GELATINOSA
ANTERIOR MEDIAN FISSURE
POSTERIOR MEDIAN SEPTUM
ANTERIOR WHITE COLUMN
POSTERIOR WHITE COLUMN
LATERAL WHITE COLUMN
ANTERIOR ROOT
POSTERIOR ROOT
SPINAL GANGLION
SPINAL NERVE

FIG. 45, 1

A spinal segment from the thoracic region viewed from the left side. (After *Gray's Anatomy*, 30th ed. 1949, p. 924. London: Longmans.)

Normal nerve fibres stain with osmic acid, OsO_4, which is reduced to a black product, probably a lower oxide or metallic osmium, by the reactive unsaturated fatty acids such as oleic acid found in the complex mixture of fats present in the myelin sheath. If normal fibres are treated with a mixture of osmic acid and potassium dichromate the unsaturated fatty acids are quickly oxidized by the dichromate so that they no longer react with the osmic acid which is, therefore, not reduced and no black colour is formed. In degenerating nerves, two weeks to two months after nerve section, the complex fatty material in myelin breaks down with the production of large amounts of unsaturated fatty acids which are not all oxidized by dichromate treatment so that the application of osmic acid simultaneously with potassium dichromate results in the formation of a black colour owing to the reduction of the osmic acid by such fatty material as has escaped oxidation by the dichromate. The degenerating nerve fibres are, therefore, stained black (Marchi's method). This technique does not show the non-myelinated components of nerves and thus can give only the position of the tracts and not the cells upon which the fibres end. The identification of the latter requires the staining of the terminal synaptic knobs (Fig. 41, 14) or their immediate preterminal fibres. Synaptic knobs begin to swell and disintegrate a day or so after they are cut off from the nerve cell; the swelling and fragmentation reach a maximum in five to ten days and disappear slowly afterwards. The origins of nerve fibres can be inferred from the disorganization of the cell body and nucleus (chromatolysis, p. 765), which occurs when the axon is cut.

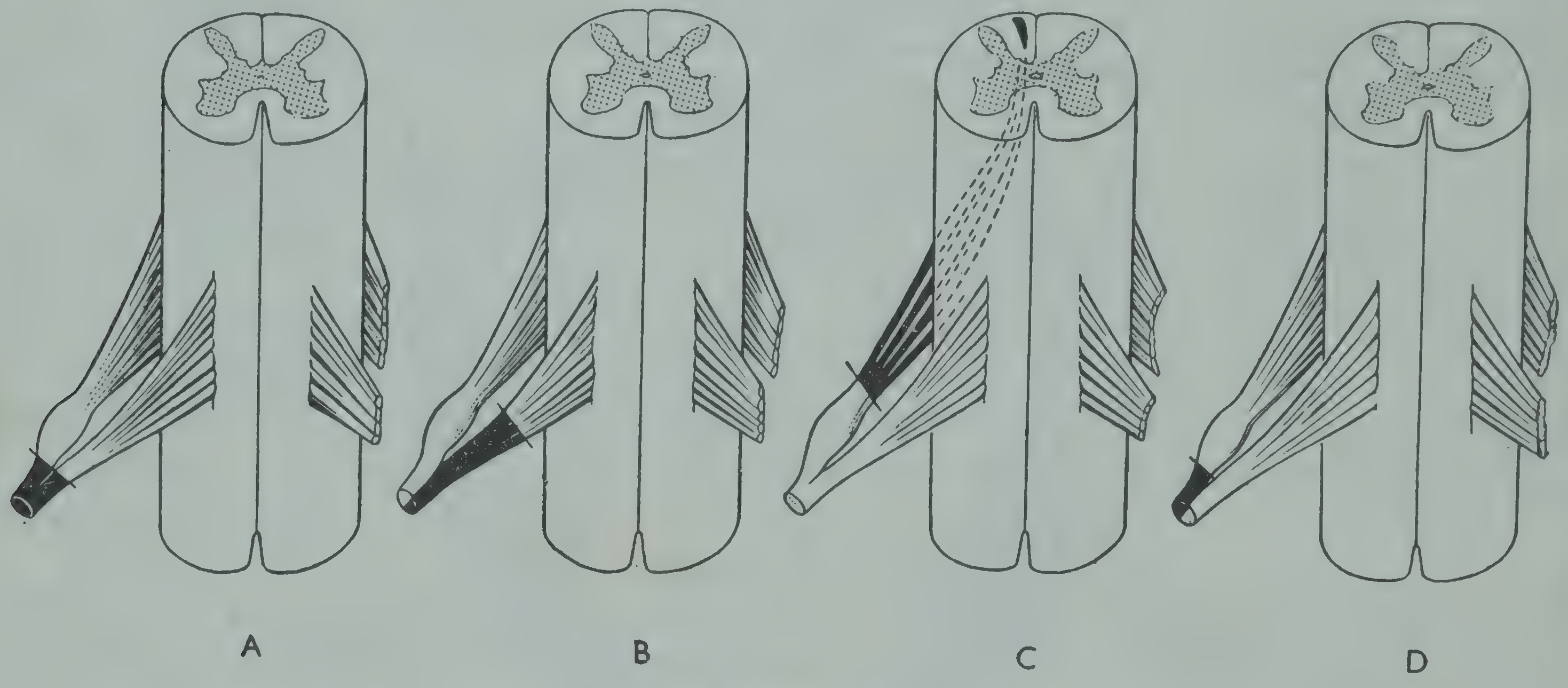


FIG. 45, 2

Diagrams to show the site of degeneration (black) after division of the spinal nerve or its roots. A, division of spinal nerve below ganglion. B, division of anterior root. C, division of posterior root above ganglion. D, division of posterior root below ganglion. (*After* G. F. Yeo (1893), *A Manual of Physiology*, p. 478. London: Churchill.)

After three months, when degeneration is complete, Marchi's method is no longer applicable and the Weigert-Pal method is employed. The tissue is first stained black with hæmatoxylin and then treated with permanganate which takes the colour out of the degenerated areas but leaves the normal fibres a deep blue-black colour.

Further information about the connexions of nerve fibres is given by the study of transneuronal degeneration; a neurone may show chromatolysis and degenerative changes if the nerve fibres sending impulses to it have degenerated, but the only good example of this is the degeneration of certain cells in the lateral geniculate body after section of one optic nerve. Another clue to their grouping is furnished by observing the myelination of nerve fibres in the embryo, since different groups acquire their myelin sheaths at different times (p. 889).

The sites of degeneration after section of the spinal nerves or their roots are shown in Fig. 45, 2. Study of the histological effects following such divisions allows of a number of conclusions. Axons from motor cells situated in the anterior horn of the grey matter run in the anterior root. The fibres of the posterior root either pass a short distance into the cord to

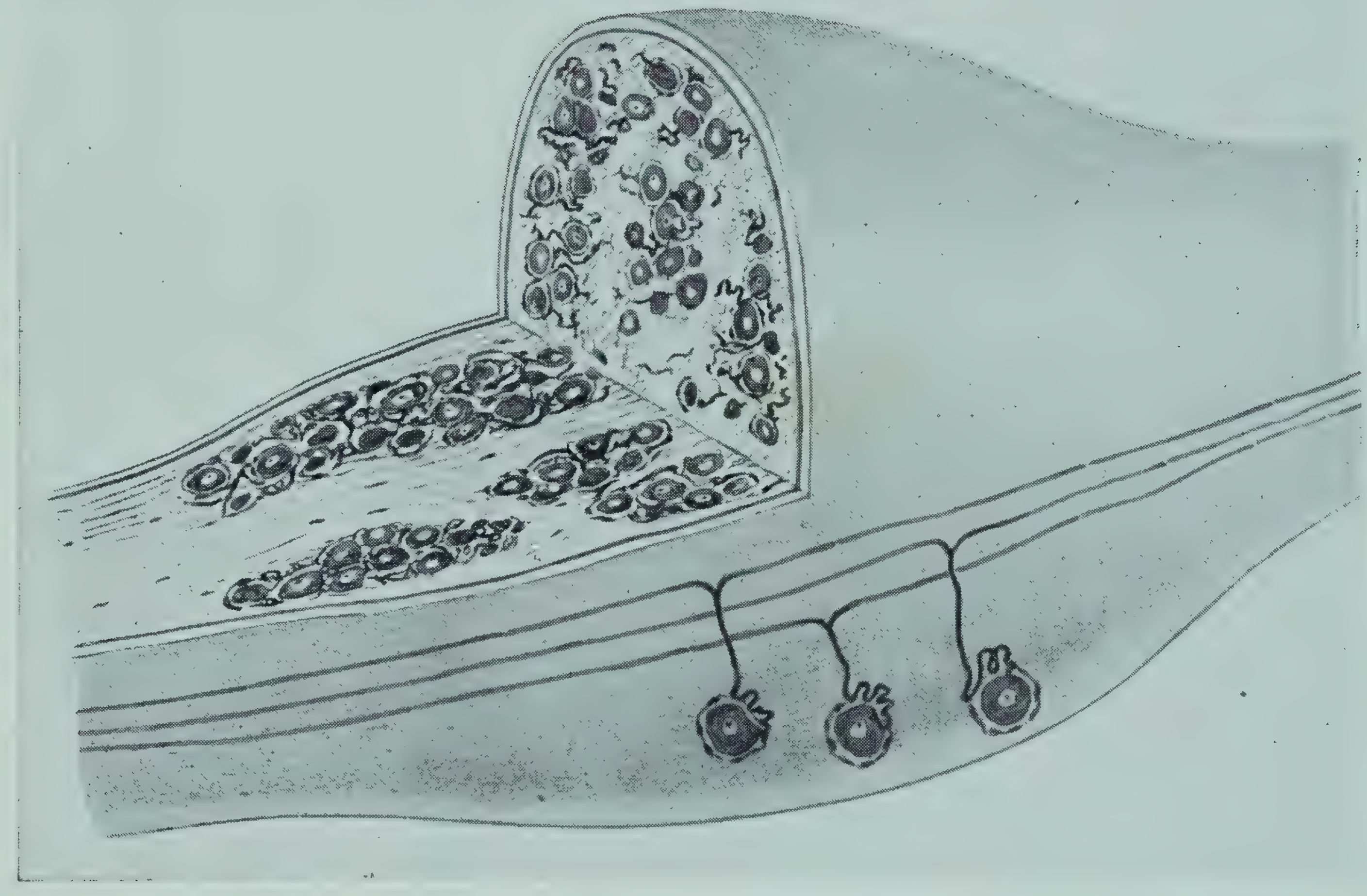


Fig. 45, 3

Diagram of spinal root ganglion.
(*After* H. S. D. Garven from Cajal and Carleton.)

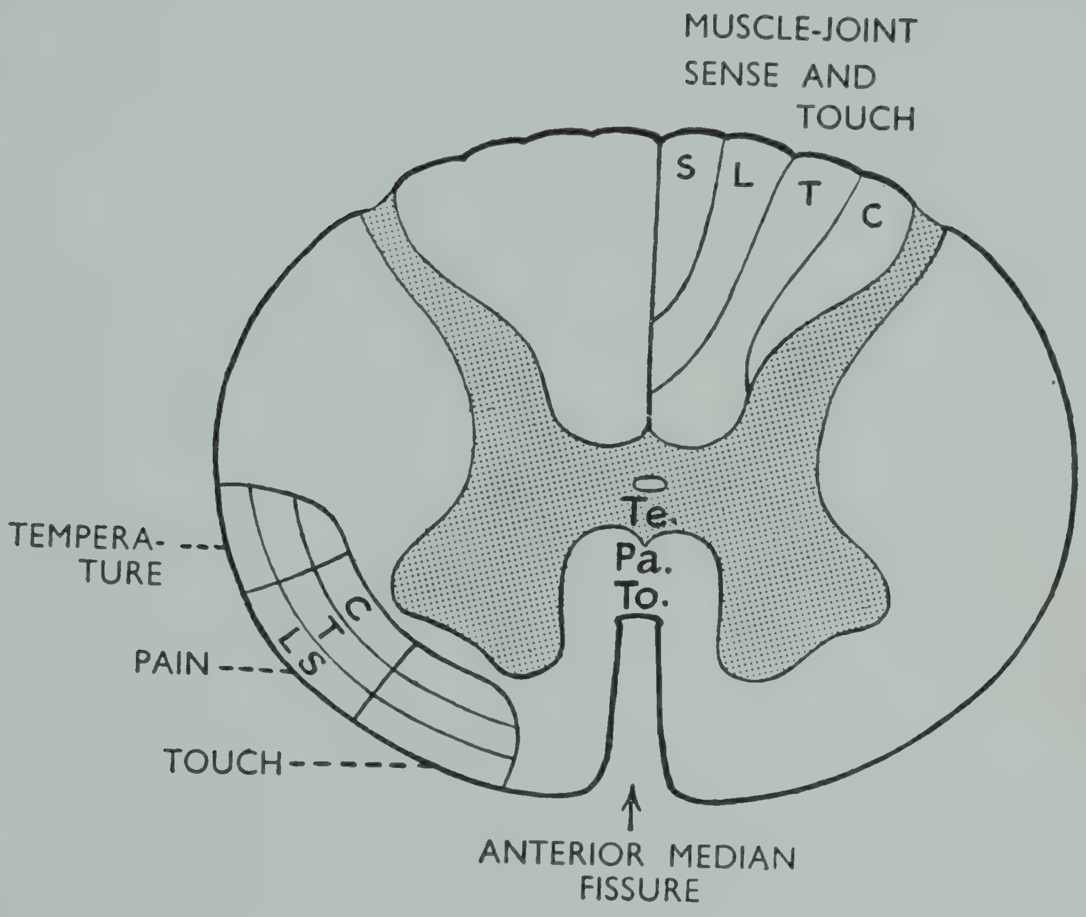

MUSCLE-JOINT SENSE AND TOUCH
S
L
T
C
TEMPERA-TURE
PAIN
TOUCH
C
T
L S
Te.
Pa.
To.
ANTERIOR MEDIAN FISSURE

Fig. 45, 4

To illustrate the positions in a cross-section of the cervical spinal cord of the chief sensory paths. The letters indicate the relative positions of the fibres carrying impulses from C, the cervical; T, thoracic; L, lumbar; S, sacral root areas. Te, Pa, To, indicate place of decussation of the temperature, pain and touch fibres. Some of the touch fibres may decussate just posterior to the central canal. (*After* Foerster.)

terminate around cells in the posterior horn, or pass up the cord in the posterior columns to the caudal part of the medulla oblongata where they end in the gracile and cuneate nuclei. The cell bodies of these posterior root fibres lie in the posterior root (spinal) ganglion.

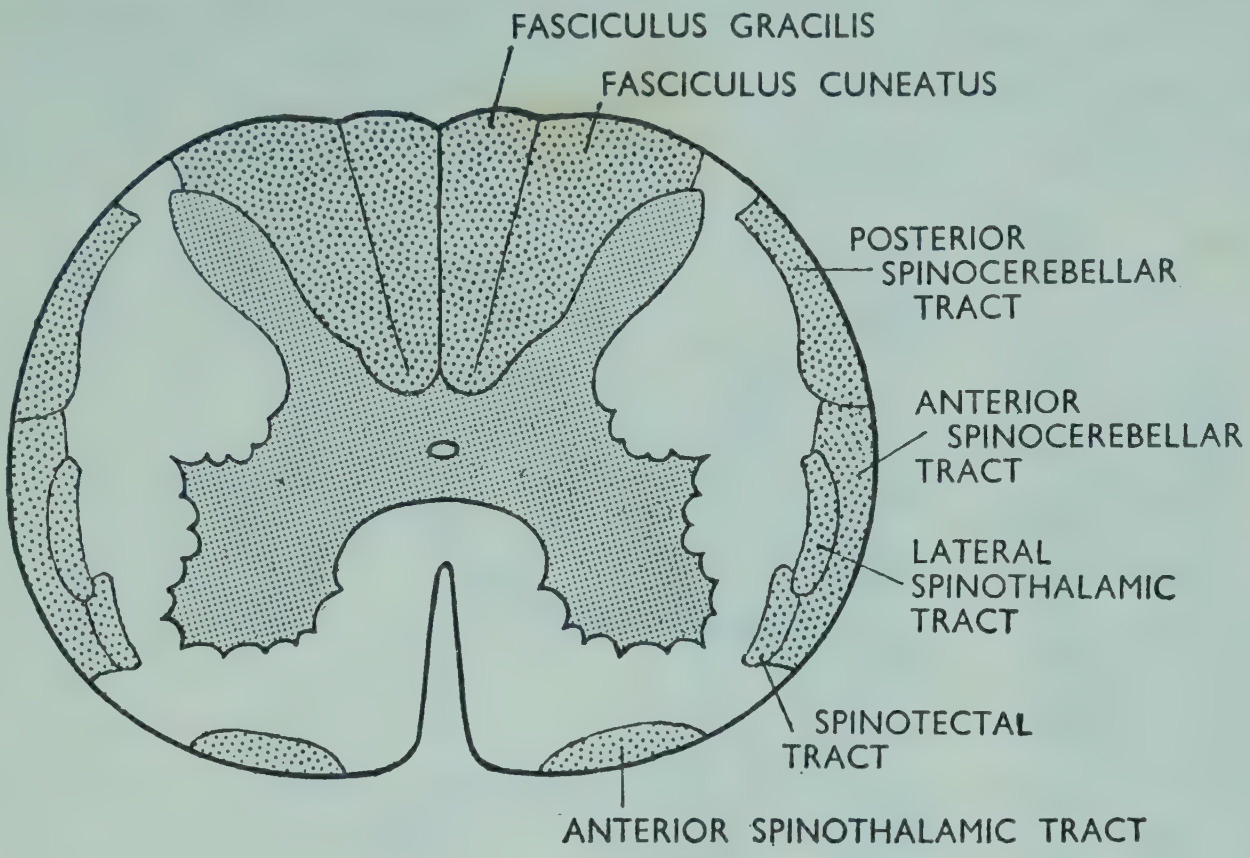

FASCICULUS GRACILIS
FASCICULUS CUNEATUS
POSTERIOR SPINOCEREBELLAR TRACT
ANTERIOR SPINOCEREBELLAR TRACT
LATERAL SPINOTHALAMIC TRACT
SPINOTECTAL TRACT
ANTERIOR SPINOTHALAMIC TRACT

Fig. 45, 5

Ascending degeneration showing the position of the ascending tracts in the human spinal cord. This diagram gives the usual view of the position of the spinothalamic tracts but their position in the anterolateral part of the cord may be more correctly given in Fig. 45, 4. Magnification about × 7.

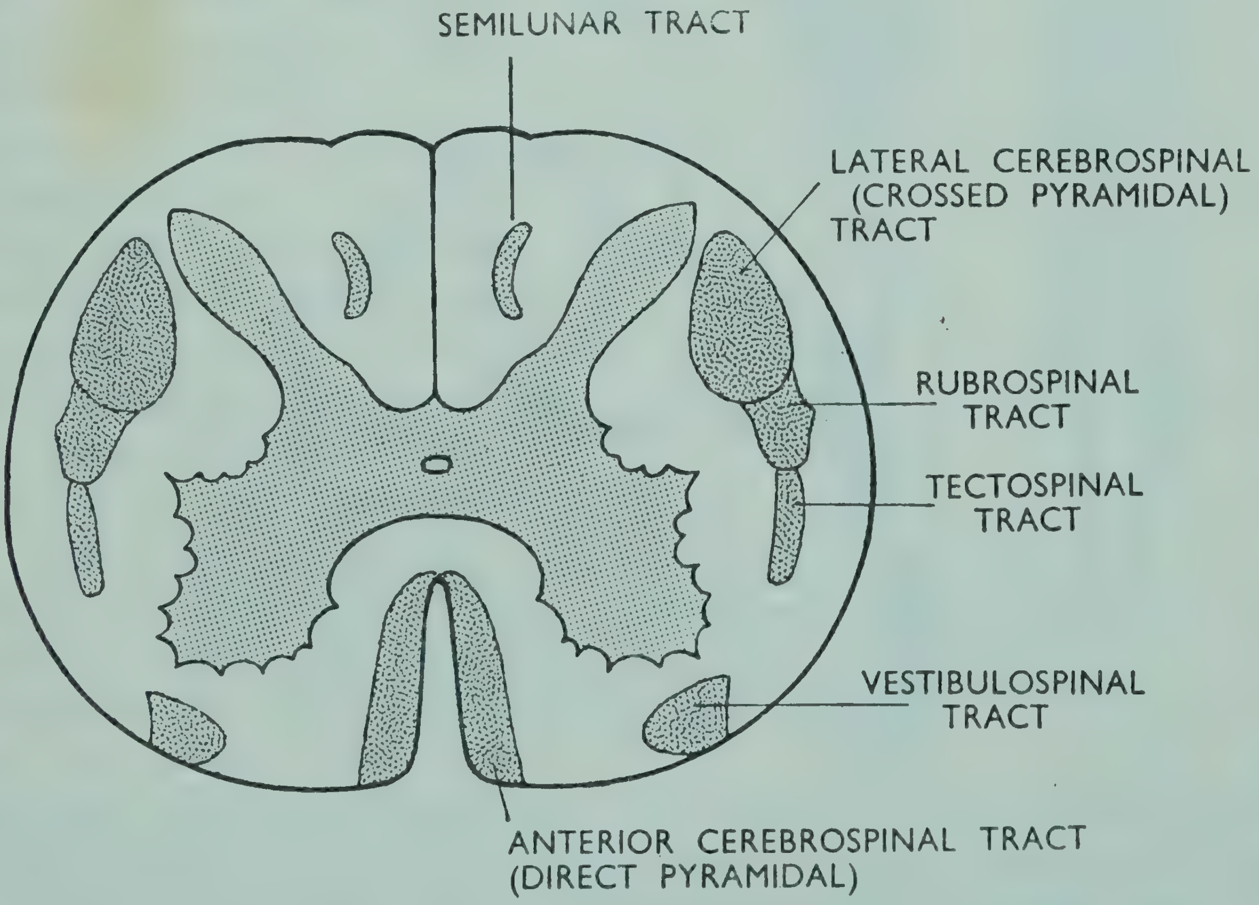

SEMILUNAR TRACT
LATERAL CEREBROSPINAL (CROSSED PYRAMIDAL) TRACT
RUBROSPINAL TRACT
TECTOSPINAL TRACT
VESTIBULOSPINAL TRACT
ANTERIOR CEREBROSPINAL TRACT (DIRECT PYRAMIDAL)

Fig. 45, 6

Descending degeneration showing the position of the descending tracts in the human spinal cord. Intersegmental tracts lie between the grey matter and the descending fibres; the semilunar tract is also intersegmental. Magnification about × 7.

When the spinal cord is cut across, degeneration occurs above (Fig. 45, 5) and below the line of section (Fig. 45, 6). That occurring below the section, *descending degeneration*, shows the position of the fibres whose cell bodies are above, that is cranial to, the cut. Conversely the degeneration occurring above the cut, *ascending degeneration*, shows the position of fibres ascending in the cord from cell bodies below the plane of section.

When the posterior root fibres are followed into the spinal ganglion it is found that the fibres have a side branch connecting them with the nerve cells in the ganglion (Fig. 45, 3). Counts of microscopic preparations show that the number of cells is very nearly (97 per cent.) the same as that of fibres. Presumably, therefore, nearly every fibre has a cell body in the ganglion. As each fibre enters the cord it divides into ascending and descending branches, each branch sending off numerous side branches at different levels in the cord (Fig. 45, 7). The posterior root fibres have three main methods of termination: (1) a very few pass to the cells of the anterior horn of grey matter to form a synapse directly with the motor neurones of the same side, the monosynaptic pathway, (2) a large number form synapses with cells at various positions in the grey matter of the posterior horn such as the cells of the nucleus proprius, the thoracic

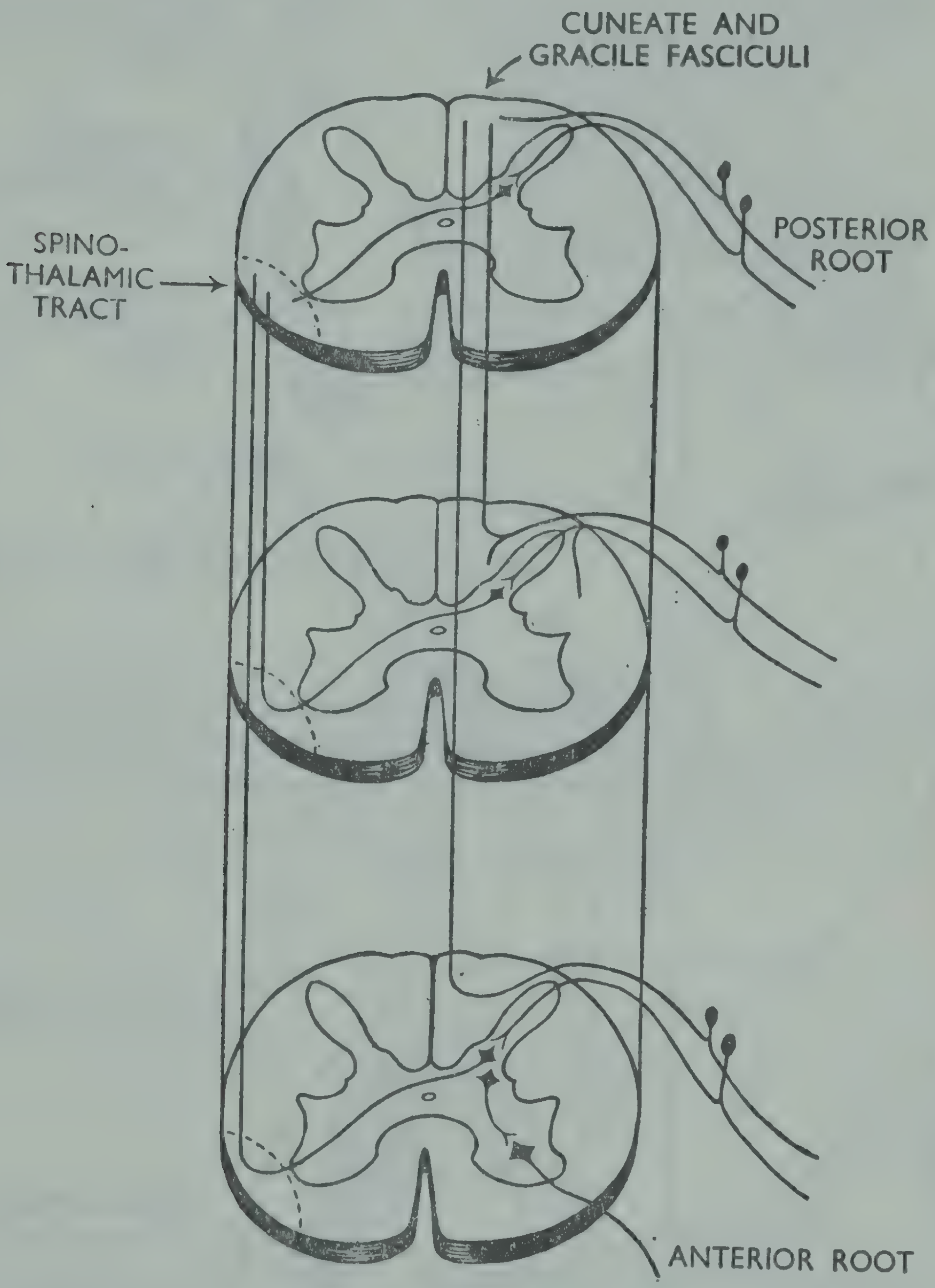


Fig. 45, 7

Diagrammatic transparency of the spinal cord viewed from the anterior (or ventral) aspect to show the paths which conduct sensation through the cord (Sir Gordon Holmes (1946), *Introduction to Clinical Neurology*, p. 66. Edinburgh: Livingstone).

nucleus and the substantia gelatinosa, and from these secondary neurones impulses pass on to other parts of the central nervous system, (3) a large number of fibres pass up the same side in the posterior columns of the white matter of the cord to the cuneate and gracile nuclei in the medulla, sending collateral branches to the grey matter as they ascend. This group contains fibres conveying impulses from the muscles (muscle-joint sense) and about half of the fibres subserving light touch.

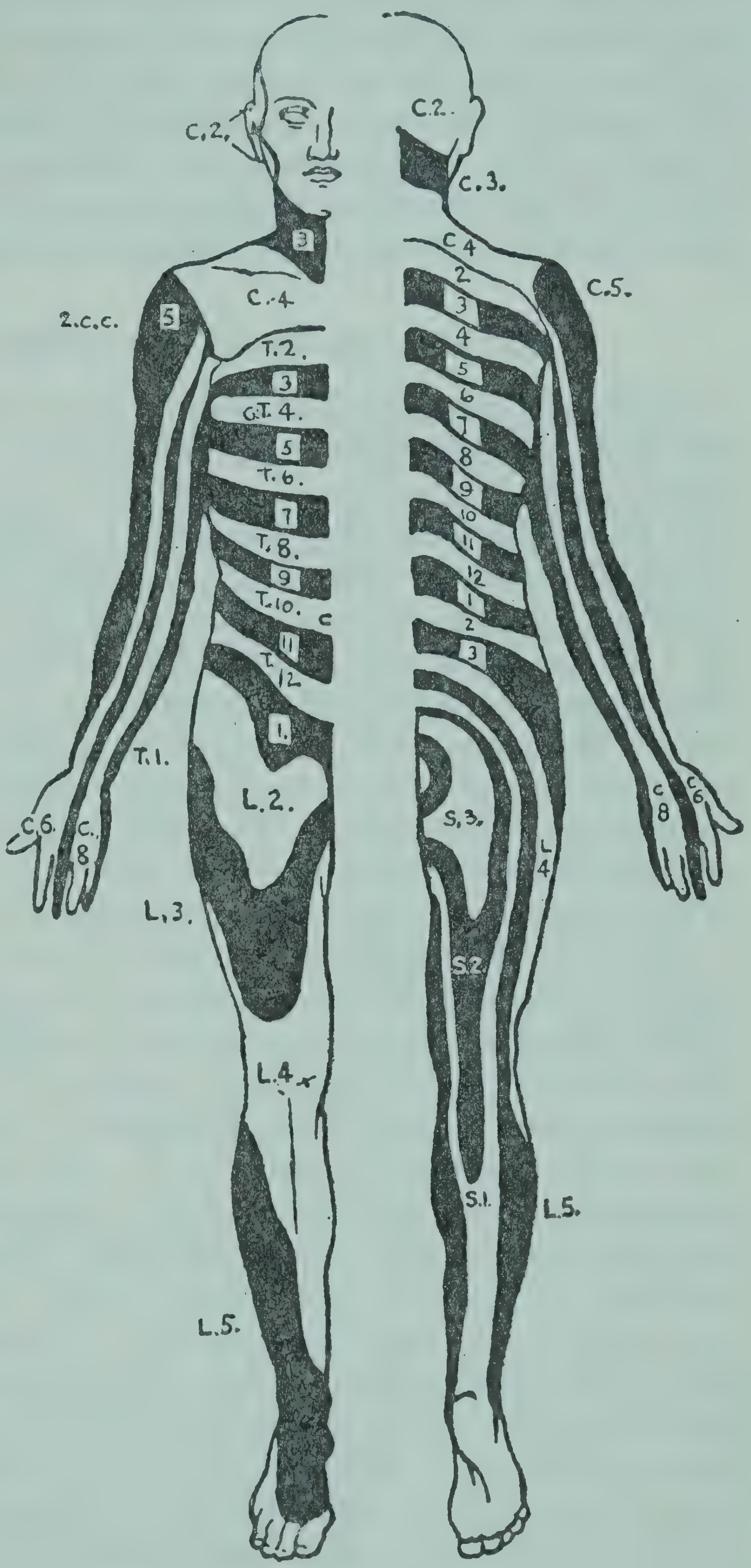


Fig. 45, 8

The cutaneous areas supplied by the posterior nerve roots (dermatomes) (Sir J. Conybeare & W. N. Mann) (1954), *Textbook of Medicine*, 11th ed. p. 70. Edinburgh: Livingstone).

The secondary neurones of group (2) have various destinations. Many are simply short interconnecting or internuncial neurones (interneurones) carrying impulses to the motoneurones of the anterior horn of grey matter of the same or adjacent segments of the cord, but other secondary neurones have much longer paths. From the thoracic nucleus (Fig. 45, 1) fibres pass to the posterior spinocerebellar tract of the same side; from other cells in the posterior horn fibres go up the anterior spinocerebellar tracts of both sides (Fig. 45, 5) carrying impulses from the limb muscles to the cerebellar cortex (mainly ipsilateral). From the nucleus proprius and perhaps also from the substantia gelatinosa the secondary neurones cross the mid-line, passing anteriorly or posteriorly to the spinal canal, and ascend in the contralateral cord as the spinothalamic tract (Fig. 45, 5), the posterior part of which carries mainly the fibres of pain and temperature while the anterior part contains chiefly touch fibres (Fig. 45, 4).

Experimental Methods of Determining Pathways

The classical method of investigating the physiology of the nervous system, namely by section and electrical stimulation, continues to yield useful information. The study of the effects of disease or of damage, due either to accident or to surgical intervention, contributes greatly to our knowledge of the central nervous system in man. Thirdly, impulses in animals can now be detected and localized in the central nervous system by micro-electrodes attached to sensitive valve amplifiers; in this way it has been possible to map out pathways through the brain of nerve impulses started off by stimuli such as sound or touch or artificial stimuli such as electrical pulses or local application of strychnine. Examples of such investigations are given in this and subsequent chapters.

Effects of Section of Roots of Spinal Nerves

For the localization of lesions of the spinal cord it is important to know the area of distribution of the posterior root fibres. The area of skin innervated by each root is called a *dermatome.* When a single posterior root is cut the area of anæsthesia is small because the skin-fields of neighbouring posterior roots overlap considerably, any one patch of skin being supplied by two, or even three, posterior roots. To demonstrate the area of a dermatome the method of *remaining sensibility* was first used by Sherrington (1893) in the monkey and much later by Foerster (1933) in man. This consists of cutting three posterior roots above and three posterior roots below the root to be studied. This root, isolated as it were from its neighbours, is left intact. The skin area which remains sensitive to mechanical stimulation is then marked out. The results of numerous experiments of this kind can be plotted as a map of the dermatomes (Fig. 45, 8).

Section of a series of posterior roots, or their destruction by disease, causes, in addition to the cutaneous anæsthesia just described, complete loss of all forms of sensation whether cutaneous, proprioceptive or visceral. Since the ingoing part of the reflex arc is interrupted, spinal reflexes are abolished and muscle tone is reduced. Loss of the protective pain reflexes frequently results in damage to the skin which may be followed by ulceration. The exclusion of the proprioceptive impulses from the cord prevents the patient from knowing, unless visually, the position of the muscles and joints in the affected area. As a result voluntary movements are carried out most inaccurately. Section of posterior roots also causes loss of vibration sense. When a heavy tuning fork is struck and laid with its base on the skin, especially over a bony prominence, the normal subject has a sensation of vibration which is conveyed by impulses passing along the afferent nerves and their continuations up the posterior columns of the cord. Vibration sense is also impaired in thalamic lesions.

Section of the anterior roots causes paralysis of the muscles supplied, with complete loss of voluntary and reflex movements. The affected muscles become flaccid, toneless and waste soon after. The cells of the anterior horn of the grey matter, the motor cells, show chromatolysis (see p. 765) which proves that they are the cell bodies of the anterior root fibres. Destruction of these cells as in anterior poliomyelitis (infantile paralysis) has the same effect as section of the anterior roots and the results, described as a lower motor neurone paralysis, are summarized in Table 45, 9. As the sympathetic fibres also emerge in the anterior roots (p. 805) their section causes a permanent loss of sweating in the area and, initially, a dilatation of the skin vessels which gradually diminishes.

Spinal Shock

After spinal transection in animals there is a period of depression of reflexes in the cord posterior (caudal) to the cut, those in the cranial part of the nervous system remaining unaffected. In the frog the spinal reflexes recover after two or three minutes and in the cat spinal reflexes may be depressed for an hour or more. In monkeys, however, the depression lasts much longer and in the chimpanzee complete depression of reflexes lasts as long as in man, that is, for several weeks. This state of depression accompanied by complete flaccidity of the muscles has been named *spinal shock*. It is not due directly to the trauma of section, since the area just cranial to the section is not depressed although the whole of the cord caudal to the section, including the most distant parts, is completely out of action. Spinal section causes a fall of blood pressure (p. 507) but this is not the cause of spinal shock since it does not appear when the blood pressure is lowered by other procedures. The prolonged nature of the so-called shock together with the slow return of tone and reflex activity suggests that the syndrome is in some way the result of the sudden interruption of impulses from higher, possibly cranial, parts of the nervous system. In the cat the state can be produced by a partial section of the ventral part of the spinal cord which divides, among others, the vestibulospinal pathways; to produce spinal shock in primates it is necessary to section the cortico-spinal (pyramidal) tracts.

Table 45, 9.—Organic Motor Paralysis

Upper Motor (corticospinal) Neurone. Supranuclear Paralysis.	*Lower Motor (nucleomuscular) Neurone. Nuclear and Infranuclear Paralysis.*
1. Movements of muscle-groups affected, never individual muscles.	1. Individual muscles may be affected.
2. Spasticity or hypertonicity of paralysed muscles.	2. Flaccidity or atonicity of paralysed muscles.
3. May have superadded 'associated movements' (synkinesiæ) on attempted voluntary movement.	3. No 'associated movements.'
4. No muscular atrophy, except from disuse.	4. Atrophy of paralysed muscles.
5. Electrical reactions normal.	5. Responses diminished or reactions of denervation (degeneration).
6. Tendon reflexes in paralysed limbs present and usually increased.	6. Tendon reflexes of paralysed muscles diminished and often absent.
7. If foot affected, plantar reflex extensor in type.	7. Plantar reflex, if present, is of normal flexor type.

(J. Purves-Stewart & C. Worster-Drought (1952), *The Diagnosis of Nervous Diseases*, 10th ed. p. 410. London: Arnold.) This is a clinical classification to be read in conjunction with p. 893 *et seq.*

Immediately after a complete transection of the spinal cord in the thoracic region in man there is complete loss of voluntary movement in both legs (*paraplegia*), with loss of sensation over the lower part of the trunk and legs. At first the leg muscles are quite flaccid, and no reflexes can be elicited. The earliest signs of return of reflex activity appear in one to six weeks, and this is accompanied by the development of flexor activity, first in the distal parts of the limbs and later in the proximal parts. The final outcome in cases of uncomplicated complete transection of the cord is chiefly extensor activity below the line of section. Extensor activity is most readily excited by proprioceptive stimuli; for example a sudden stretch of the iliopsoas. The knee jerk and other reflexes are easily elicited and are exaggerated. Stroking the sole of the foot readily produces a dorsiflexion (an upturning) of the great toe (Babinski response). Reflex contraction of the large bowel and of the urinary bladder often occur simultaneously and some patients are able to regulate their bowel movements by inserting a suppository into the rectum or even by digital stretching of the anal sphincter. Mass reflexes involving simultaneous, violent contractions of the limbs, bladder and bowel, accompanied by sweating, were described after the 1914–1918 war in spinal man following gun-shot wounds. They may be regarded as an abnormal spread of reflex activity, related in some way to infection of bladder or of bed sores, both of which can now be avoided by appropriate precautions. Further details of the reflexes of spinal man are given on pp. 278, 627 and 815.

Hemisection of Spinal Cord

A hemisection produces paralysis of movement with loss of position sense on the same side below the level of the cut and on the opposite side loss of sensation (anæsthesia) of pain and temperature without loss of movement. The reason for this characteristic distribution of motor and sensory loss is clear from Fig. 45, 10. The band of hyperæsthesia on the same side is usually attributed to irritation of the neurones adjacent to the lesion. If the local lesion destroys at least three segments (see dermatomes, p. 827) there may be a band of impaired sensation in the dermatomes corresponding to the level of the lesion. When the anterior horn cells are involved a paralysis of the lower motor neurone type affects the muscles innervated by the corresponding segments. Since the first account of the effects of hemisection was given by Brown-Séquard the syndrome is now described by his name.

Disease causes incomplete transections of the cord much more commonly than complete transections. An important characteristic of incomplete sections is the tendency to spasm of the extensor muscles. In animals extensor spasms occur when even a small portion of the corticospinal (pyramidal) tracts or the vestibulospinal tracts is left intact, but when both are destroyed extensor reflexes disappear.

PROPERTIES OF SPINAL REFLEXES

Flexion Reflexes

If the foot of a decapitated frog is pinched the legs move quickly away from the harmful stimulus but, if the spinal cord of this brainless preparation is destroyed by pithing, reflex withdrawal no longer occurs. This funda-

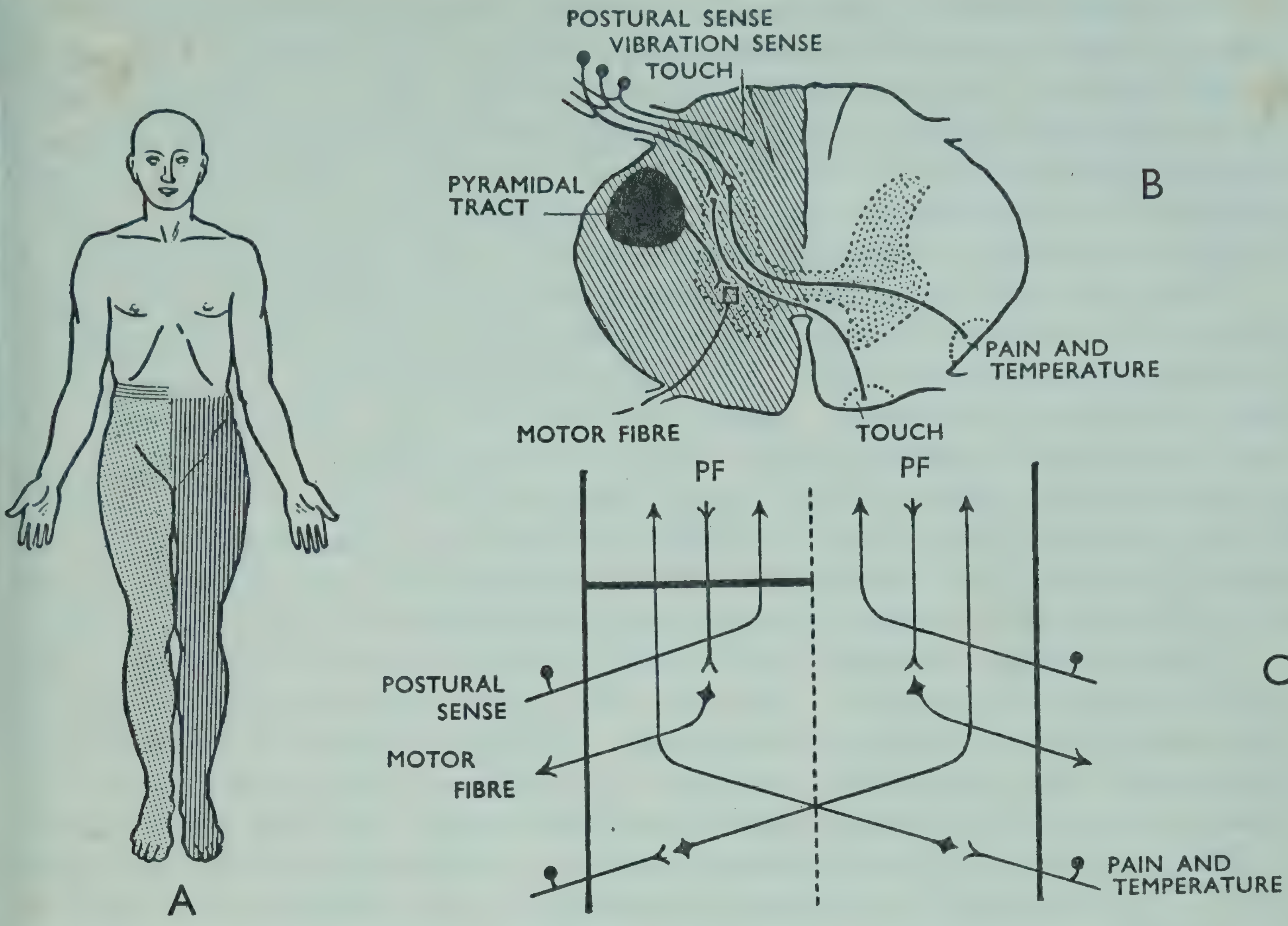


FIG. 45, 10

THE BROWN-SÉQUARD SYNDROME

Illustrating the Brown-Séquard syndrome from a lesion of the right half of the cord. In A the distribution of cutaneous sensory change is indicated by horizontal and vertical shading. The area of spastic paresis is indicated by stippling. In B, the shaded (right) half of the cord is the seat of the lesion and the diagram indicates the pathways interrupted. A diagrammatic longitudinal section is given in C. The black line indicates the site of the hemisection. The ascending and descending tracts which are interrupted are labelled. PF, pyramidal fibre. The resulting disorders are given in the table below.

ON THE SIDE OF THE LESION	ON THE OPPOSITE SIDE
Sensory	
At level of lesion.—A band of cutaneous hyperæsthesia may be present. Alternatively and only if three successive spinal segments are involved, there may be a band of impairment of cutaneous sensibility to touch, pain and temperature.	*At level of lesion.*—No abnormality.
Below level of lesion.—Tactile, painful and thermal sensibility normal. Postural and vibratory sensibility (p. 828) impaired or lost.	*Below level of lesion.*—Impairment or loss of painful and thermal sensibility. Tactile, postural and vibratory sensibility normal.
Motor	
Below level of lesion.—Spastic paresis of leg, increased tendon jerks, clonus, extensor plantar response.	*Below level of lesion.*—No abnormality.

(Sir Francis Walshe (1955), *Diseases of the Nervous System*, 8th ed. p. 38. Edinburgh: Livingstone.)

mental observation was made by Stephen Hales about 1730 but the word 'reflex' was not used until about forty years later, and serious investigation was delayed till the end of last century when Sherrington took up the subject.

Although it is theoretically possible for impulses coming in at a posterior root to cause motor impulses to leave in the anterior root at the same level and so produce muscular movements, it is unusual for a reflex to be confined to one segment. In the withdrawal reflex in the frog just described there is a co-ordinated flexion of the ankle, knee and hip by synergistic groups of muscles innervated by a number of spinal segments. The idea of a spinal segment is convenient for anatomical description but a segment is not a functional unit as even simple reflexes involve many segments of the cord.

Since the activities of the spinal cord are greatly modified by the higher parts of the nervous system it is necessary for analysis of spinal reflexes to make a *spinal preparation*. The cat and dog have most frequently been used in such experiments. After the cord has been divided in the neck there is a period of spinal shock (p. 829), but if respiration is maintained by a pump the shock passes off and many reflexes can be elicited. The activity of the muscles to be studied is usually recorded by attaching them to isometric levers and the conditions of the experiment may be made as simple as possible by denervation of neighbouring structures or by removal of skin.

If an injurious stimulus, one which would be painful if applied to the intact animal, is applied to the foot of a spinal preparation the flexion reflex results in a withdrawal of the foot. This phenomenon is described by Sherrington as a *nociceptive response*. Withdrawal also occurs when any sensory nerve in the limb is stimulated electrically and this phenomenon is often made use of in experimental analysis of the properties of the spinal cord. For example, by varying the strength of the stimulus it is easy to demonstrate *irradiation*. A small stimulus causes a small response confined to a few muscles, whereas stronger stimulation involves many muscles in the limb and may cause movements in other limbs.

The time elapsing between the stimulus and the reflex response, the total latent period of a reflex, can easily be measured from a kymographic tracing. If the time taken for the nerve impulses to pass inward along sensory nerves and out again along the motor nerves is subtracted from the total reflex time, we have the *central reflex time*, or *reduced reflex time*, which in a spinal reflex must be due mainly to synaptic delay. For a single volley set up by a single faradic stimulus the central reflex time for a flexion reflex in the cat is about 4 msec. When two subliminal volleys reach the cord with only a short interval, say 7 msec, between them, the central reflex time for the flexion reflex evoked by the second volley is reduced, in some cases to 0·5 msec. Apparently the rise in central excitatory state (p. 747), which follows the first volley, can be quickly raised by the second volley to a level at which the motor neurones discharge. A more recent conception of what is happening is that the first volley, although not sufficient to excite the motoneurones, sets the chains of internuncial neurones (p. 753) reverberating and in this way the motor neurones are continuously bombarded with impulses which facilitate the action of the next volley. Eccles believes that each excitory impulse arriving at a motoneurone reduces the potential across the cell membrane and a rapid succession of such impulses can reduce the potential sufficiently to make the motoneurone discharge. Since the synaptic delay at the motoneurone is of the order of 0·5 msec the central reflex time can never be less than this.

When a motor nerve to a muscle is stimulated by a single electrical impulse a single volley travels down to the muscle which responds by showing a sudden abrupt rise of tension followed by a quick relaxation,

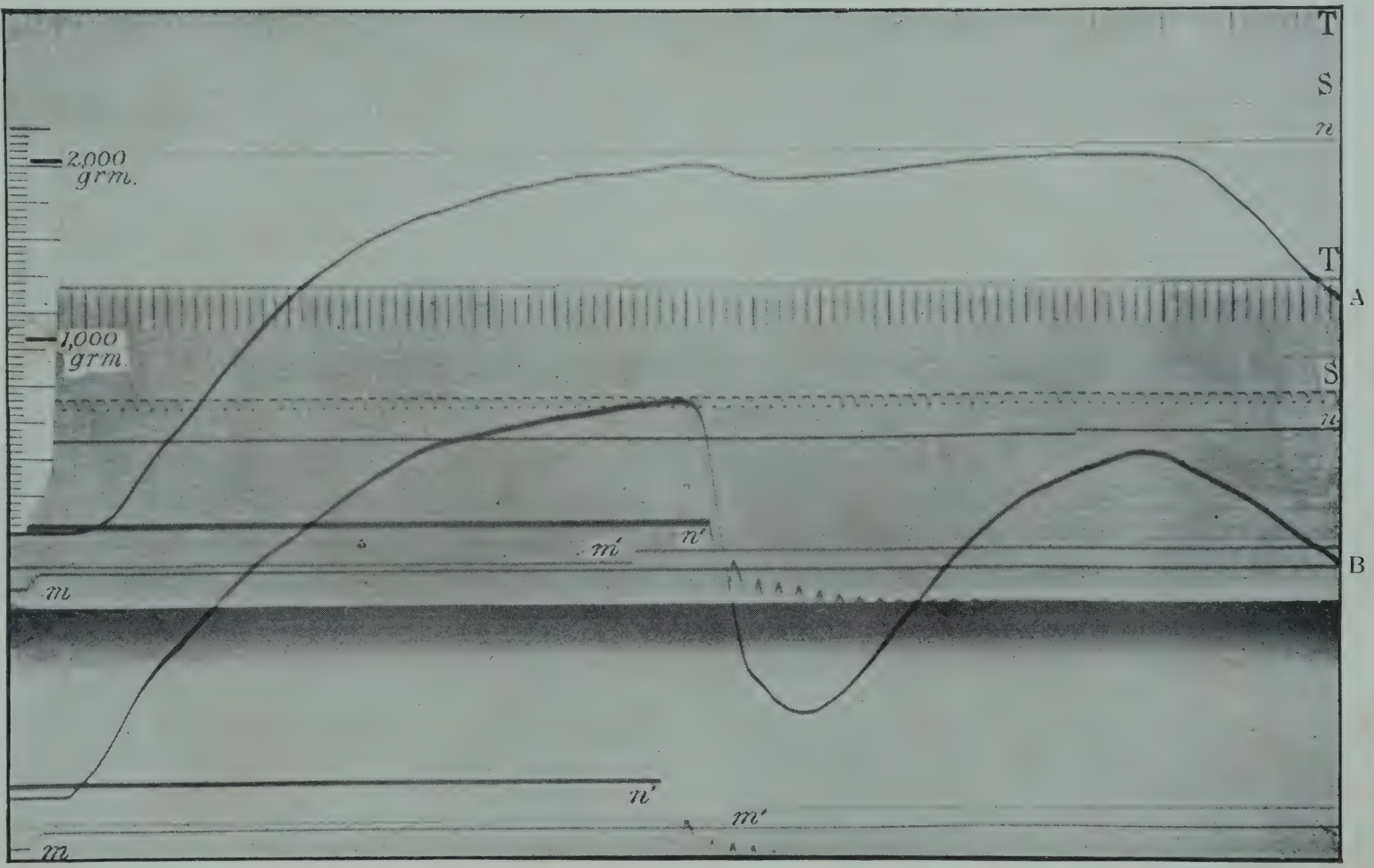


FIG. 45, 11

Records of vastocrureus muscle in a decerebrate cat. Reflex tetanus from crossed peroneo-popliteal nerve by break-shocks 48 per sec. Same intensity of stimuli for both reflexes; in the upper a lapse of six consecutive stimuli during short-circuit marked by signals *m′* and *n′*; in the lower no lapse of the stimuli to the excitatory afferent nerve, but an intercurrent faradization of the ipsilateral peroneo-popliteal nerve for the period marked between signals *n′* and *m′*, a period corresponding to the delivery of three consecutive stimuli to the contralateral afferent nerve. Time, 1/50 sec. Same tension scale applicable to both curves. Lettering as in Fig. 45, 15 (E. G. T. Liddell and C. S. Sherrington (1924), *Proc. Roy. Soc.* B, **95,** 142).

that is, a twitch. A single volley reaching the cord along an afferent nerve evokes a muscular response which is less abrupt in onset and more sustained than a twitch; the latent period of the reflex contraction is, of course, longer than the latent period when the motor nerve is stimulated. The muscle tension may rise relatively slowly if the afferent volley does not bring all the accessible motoneurones into action at the one instant. The reflex contraction lasts somewhat longer and often shows a higher tension than a twitch produced by stimulation of the motor nerve. This is due to irregular repetitive firing of the motoneurones termed *after-discharge*, which can be explained on the basis of the activity of reverberating internuncial neurones (p. 753). The after-discharge also accounts for the relatively slow relaxation at the end of a reflex contraction.

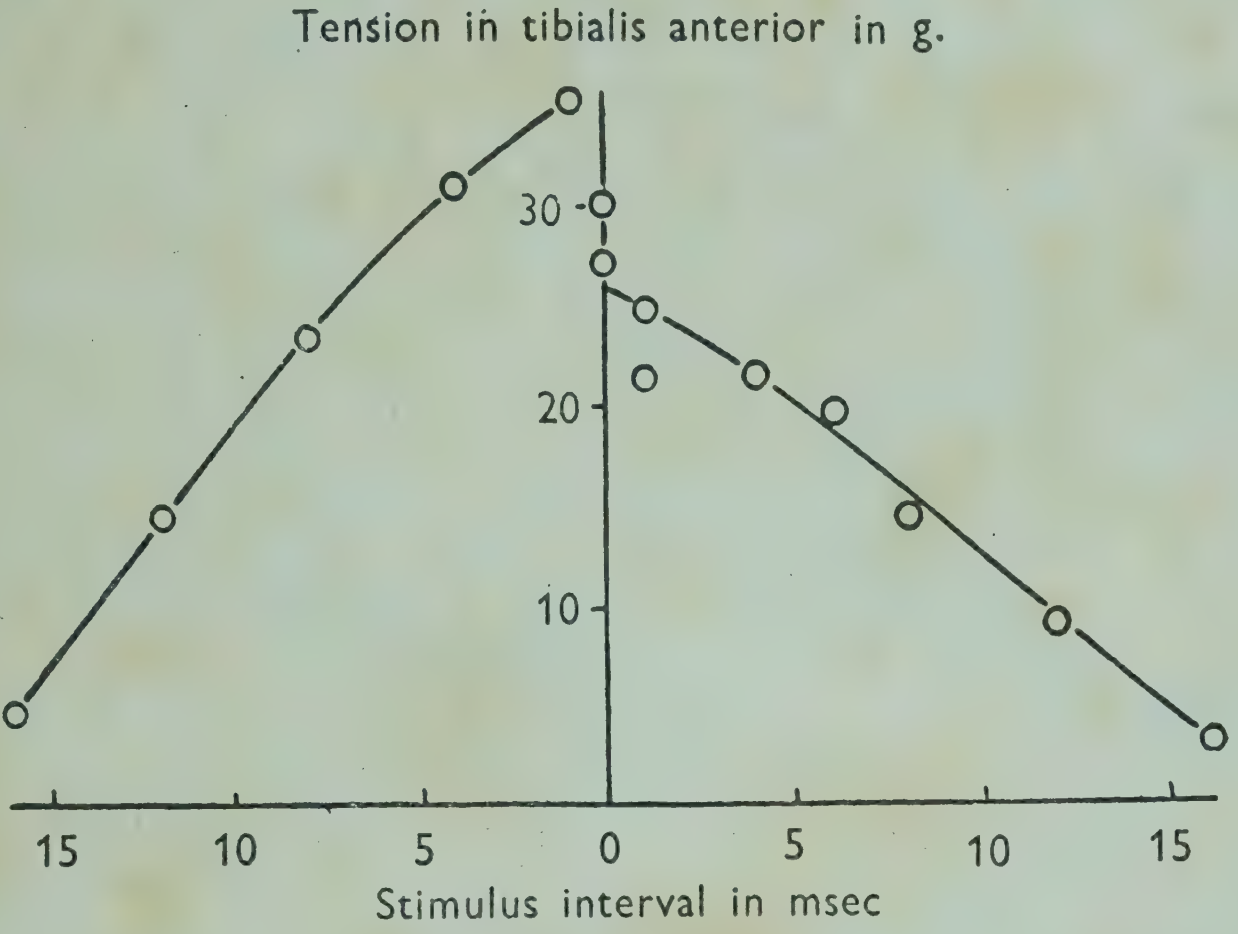
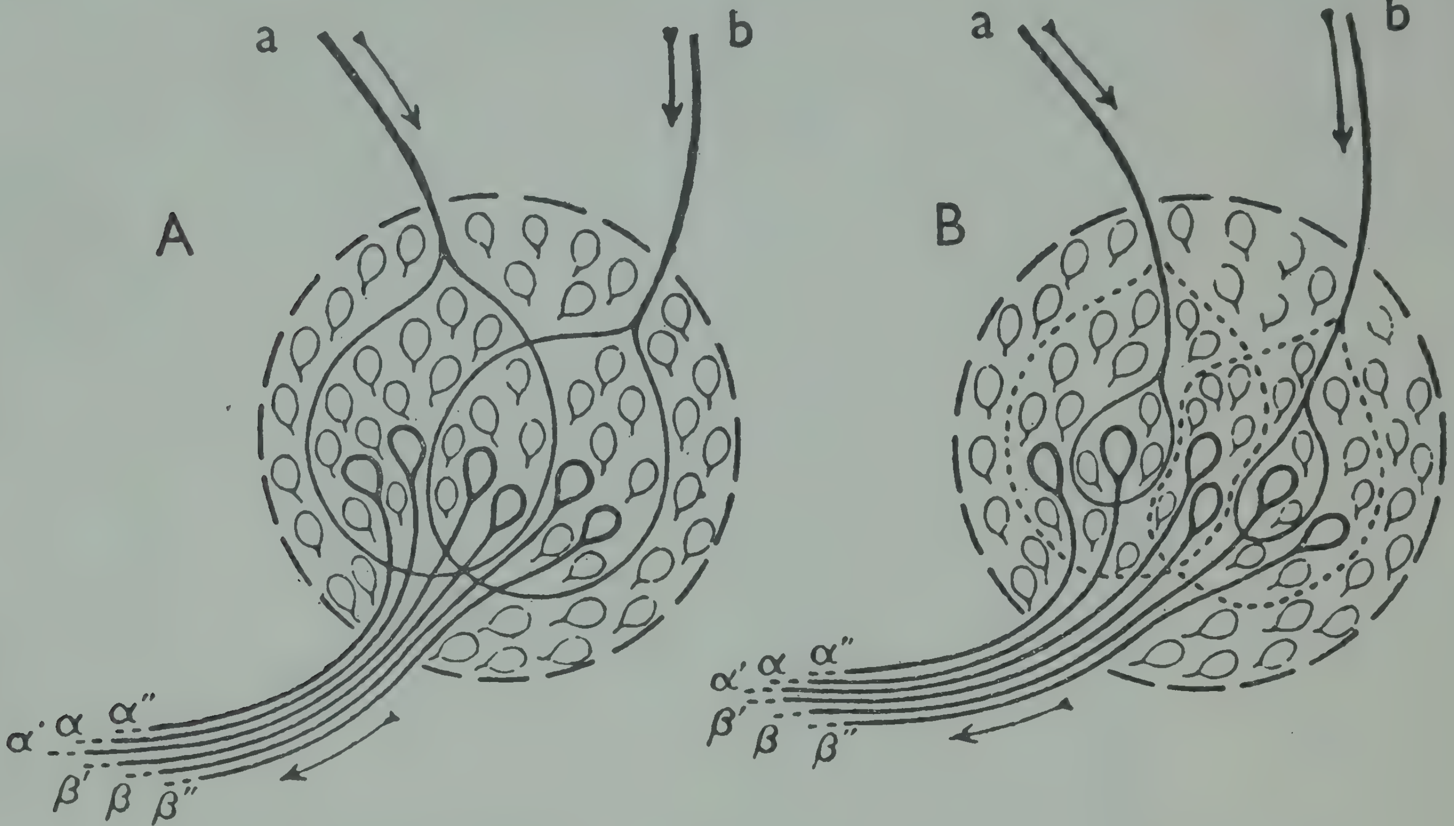


Fig. 45, 12

To illustrate spatial summation. Reflex responses of the tibialis anterior muscle in a spinal cat to two stimuli (each one of which alone is just threshold) to two separate afferent nerves in the same limb at various intervals. To right of zero, the nerve of the lateral gastrocnemius muscle is stimulated first; to left of zero, the nerve of the medial gastrocnemius muscle first. Curve shows the relation between resulting tension and the stimulus interval (R. S. Creed, D. Denny-Brown, J. C. Eccles, E. G. T. Liddell & C. S. Sherrington (1932), *Reflex Activity of the Spinal Cord*, p. 32. Oxford: Clarendon Press).

If two stimuli are applied to an afferent nerve at an interval greater than the refractory period of the nerve two reflex muscle responses are obtained, but if the separation between the stimuli is less than the refractory period (say 0·5 msec) the second is ineffective. Another consequence of the properties of peripheral nerves is that two subliminal stimuli following one another at a very short interval (up to 0·2 msec) may elicit a reflex effect. This temporal summation is of no physiological importance since the smallest interval in which it occurs is less than the refractory period of the afferent nerve. A much more important mechanism, described as *spatial summation* (simultaneous spinal induction), is shown by the following experiment (Fig. 45, 12). The tibialis anterior muscle is attached to an isometric lever and the medial and lateral branches of the sciatic nerve are dissected out and laid on stimulating electrodes. The strength of stimulus (the threshold) required to be applied to each branch to produce a just perceptible reflex increase of tension in the muscle is found. If two such stimuli are sent in simultaneously to the spinal cord by both branches of the sciatic nerve quite a considerable tension is aroused reflexly—about 30 g. in the experiment shown. If there is an interval between the stimuli the tension developed is less, but spatial summation is still detectable even when the interval between the stimuli is as long as 15 msec.



A.—Occlusion B.—Facilitation

Fig. 45, 13

A.—Two excitatory afferents, a and b, with their fields of supraliminal effect in the motoneurone pool of a muscle. a activates by itself 4 units (α', α, α'' and β'); b by itself 4 (β', β, β'' and α'). Concurrently they activate not 8 but 6, that is, give contraction-deficit by occlusion of contraction in α' and β'.

B.—Weaker stimulation of a and b restricting their supraliminal fields of effect in the pool as shown by the continuous-line limit: a by itself activates 1 unit; b similarly; concurrently they activate 4 units (α', α, β' and β) owing to summation of subliminal effect in the overlap of the subliminal fields outlined by dots. (Subliminal fields of effect are not indicated in diagram A) (C. S. Sherrington (1929), *Proc. Roy. Soc.* B, **105,** 338.)

Sherrington has supposed that the arrival of afferent impulses at the motoneurones increases the *central excitatory state* (c.e.s.) and that when this rises to a certain critical value the motoneurones discharge and a reflex contraction occurs. If the experiment just described is considered in conjunction with Fig. 45, 13, B, an explanation of spatial summation can be given. When a weak stimulus is sent in at a one motoneurone discharges and a small contraction is obtained; at the same time the excitatory state in the neighbouring motoneurones (within the dotted lines) is raised. This part of the motoneurone pool is described as the *subliminal fringe.* A similar state of affairs occurs when b is weakly stimulated but, if both a and b are stimulated at the same time, the subliminal fringes overlap and the c.e.s. in the area common to both may rise to a value at which a number of motoneurones discharge and produce a large muscular contraction.

The experiment on the flexion reflex just described indicates that the c.e.s. lasts about 15 msec. Other experiments on these lines show that c.e.s. rises fairly quickly after an afferent volley but falls away spontaneously rather more slowly. The first explanation which was suggested for the occurrence of c.e.s., namely, that it is due to accumulation of a chemical transmitter substance at the synapse is not very satisfactory and it is perhaps better to think of it as a partial depolarization of the motor neurones or as a continuous bombardment by internuncial neurones (pp. 751 and 753).

TABLE 45, 14

To show the Tension developed by the Semitendinosus Muscle of a Cat made to contract directly and reflexly

	Isometric Tension g.	Percentage of Maximum
Direct stimulation of the motor nerve to the semitendinosus muscle (maximal motor tetanus)	3000	100
Maximal stimulation of the afferent nerves named below yielded the tensions shown in the next column:		
Anterior tibial nerve	2900	97
A dorsal digital nerve	2400	80
Small sciatic nerve	1860	63
Nerve of sartorius	1500	50
External cutaneous (groin) nerve . . .	830	28
Obturator (deep) nerve	630	21

(S. Cooper, D. E. Denny-Brown & C. S. Sherrington (1926), *Proc. Roy. Soc.* B, **100**, 460.)

If an afferent nerve is stimulated by a rapid series of stimuli the flexor muscle of the limb shows a reflex tetanus. The tension developed depends on the strength of the afferent stimulation up to a maximum tension which is never as great as that shown in a maximal tetanus obtained by direct faradic stimulation of the motor nerve (Table 45, 14). In other words, no single afferent nerve can reach all the motoneurones of a given muscle; some, however, can command a larger fraction than others. This phenomenon is described as *fractionation.* The fraction is not fixed for any afferent nerve, since it is reduced in ‘spinal shock’ (p. 829) and increased if the animal is given strychnine in amounts short of that required to produce convulsions.

The results given in Table 45, 14, show that any one motor neurone is played upon by afferents arising in many different parts of the body. There is thus an overlap on the afferent side of the reflex arc caused by the convergence of many afferents on each motor neurone, which behaves as a *final common path.* This convergence also leads to *occlusion* which is explained in Fig. 45, 13, A. Strong stimulation of afferent *a* alone brings four motoneurones into action and stimulation of afferent *b* alone makes four motoneurones discharge; simultaneous stimulation of the two afferents, because of their overlap, brings only six, not eight, motoneurones into action. As would be expected, therefore, in experiments on the flexion reflex in a spinal animal it is found that the tension produced by the simultaneous maximal stimulation of two afferents is only a little greater than if either afferent is stimulated alone.

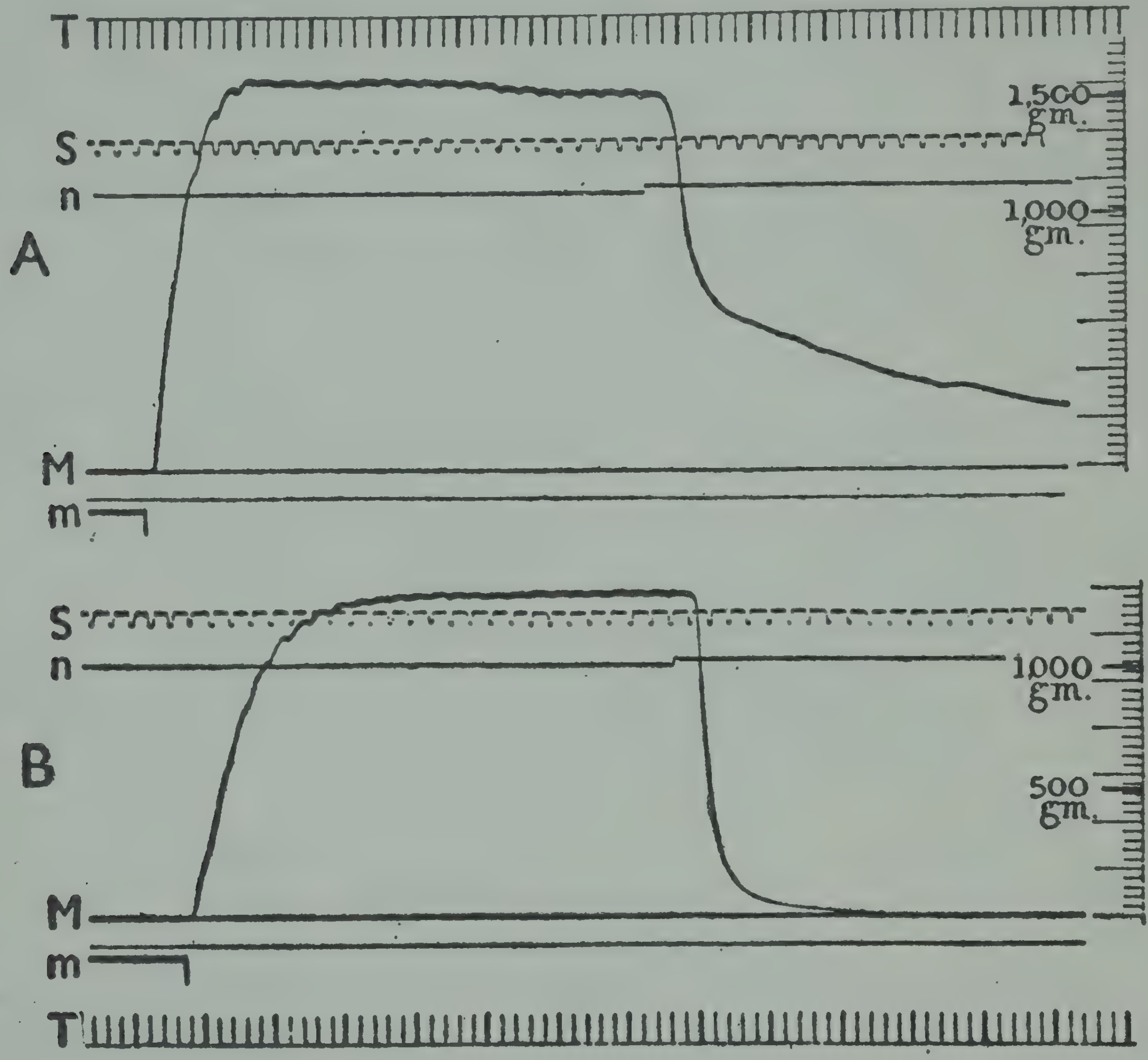


FIG. 45, 15

Record of contraction of the semitendinosus, a flexor muscle of the leg. Spinal preparation. M = myograph; *m* to *n* = unshorting and shorting of stimulating circuit; S = stimulating key; A = reflex tetanus, afferent nerve stimulated was the ipsilateral lateral popliteal; B = motor nerve tetanus, in same preparation as yielded A. In both break-shocks at 38 per sec. T, in 1/50 sec; tendon movement × 35 (E. G. T. Liddell & C. S. Sherrington (1923), *Proc. Roy. Soc.* B, **95**, 142).

It is a characteristic of the flexion reflex that the tension of the flexor muscle rises very suddenly—even more suddenly than that in a tetanus by direct stimulation of the motor nerve (Fig. 45, 15). This suggests that all the motor neurones involved discharge simultaneously (the reflex is said to be *d'emblée*) and there is no gradual *recruitment* of neurones. The latent period is very short—of the order of 10 msec. A smooth tetanus is produced only when a high rate of stimulation is used, say 40 per second. This description is very different from that of the extension reflex which is considered on the next page. (See Table 45, 16.)

If a reflex is repeatedly elicited the response of the muscle gradually declines and may eventually cease. Since direct stimulation of the motor nerve or the muscle itself is still able to produce a contraction it is evident that the *fatigue* must depend on changes in the central nervous system, probably as the result of changes at the synapses. The scratch reflex in the dog (p. 846), elicited by stimulation of the skin over the thorax, shows fatigue quite readily if the point of stimulation is kept constant but if this is changed slowly the rhythmic scratching movements can be greatly prolonged. This shows that fatigue is likely to be due to changes in the afferent side of the reflex arc. Reflexes vary very much in their liability to fatigue; the scratch reflex (p. 846) is easily fatigued but the postural reflexes (p. 855) can be maintained almost indefinitely. Fatigue can be regarded as an increase of resistance to passage of a nerve impulse across a synapse which may increase up to a complete block. It is obvious that quite apart from fatigue there must be a variation in the resistance of the pathways throughout the central nervous system, for if this were not so, reflex responses would be unlikely to follow any definite pattern and an afferent impulse, because the interconnexions within the nervous system are so great, might send enormous numbers of motor neurones into action. This effect is seen in strychnine poisoning and in tetanus (lock-jaw), where a slight touch on the skin may send the animal or man into generalized convulsions.

TABLE 45, 16

Characteristics of Flexion and Extension Reflexes

	Flexion Reflex	Extension Reflex
Latent period . .	Short (10 msec)	Long (40 to 100 msec)
Tension . . .	Rises suddenly Declines quickly	Rises slowly (recruitment) Declines slowly (after-discharge)
Stimulation for complete tetanus . . .	40 per sec to ipsilateral afferent nerve	5 per sec to contralateral afferent nerve
Central excitatory state.	Dies away quickly	Dies away slowly

EXTENSION REFLEXES

More complicated reflexes than the flexion reflex can be demonstrated in the spinal animal. If the *right* quadriceps femoris muscle (the extensor of the leg) is connected to an isometric lever, stimulation of the central end of the *left* lateral popliteal nerve produces a contraction which can be recorded. This is called a *crossed extension reflex* (Fig. 45, 17, A). The latent period is from 40 to 100 msec which is several times that of a flexor reflex. This reflex can be elicited by a single afferent volley which is followed by a succession of volleys from the motoneurones. Thus a very slow rate of stimulation of an afferent nerve, say 5 per second, results in a smooth tetanus which could be produced by direct stimulation of the motor nerve only if 40 volleys per second arrived at the muscle. As stimulation of the afferent nerve continues the tension rises slowly as the number of motoneurones involved gradually increases. This phenomenon is called *recruitment* (Fig. 45, 17). The tension does not fall off as soon as the stimulation of the afferent nerve ceases but continues for some time and then declines slowly because the motoneurones continue to discharge; after-discharge is a prominent characteristic of crossed extension reflexes. Extension reflexes differ from flexion reflexes in that in the former c.e.s. dies away spontaneously much more slowly.

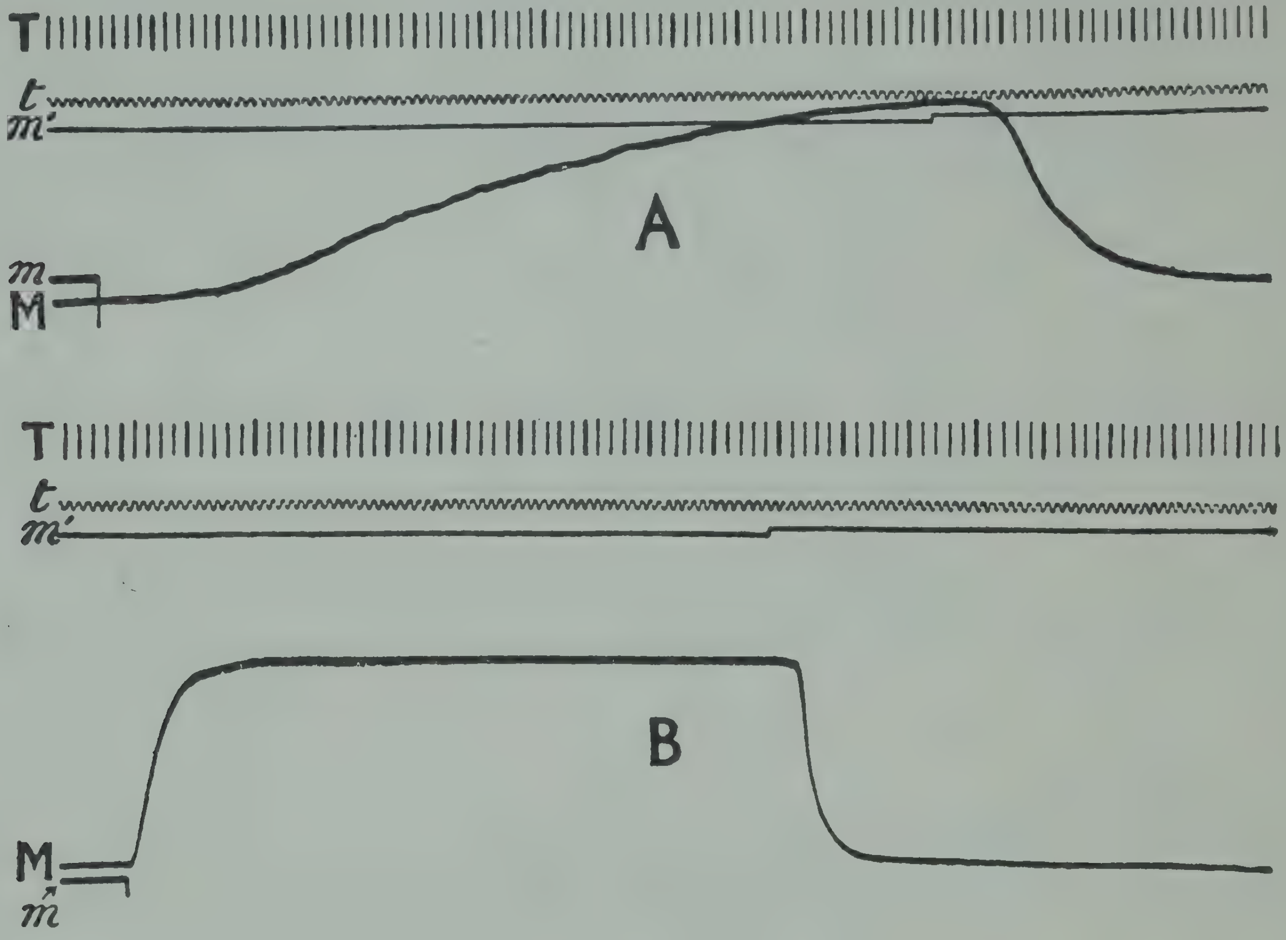


Fig. 45, 17

Record of contractions of the quadriceps extensor, an extensor muscle of the leg, of a decerebrate preparation. t = frequency of stimulus; m′ = shorting of stimulating circuit for tetanic series; m = unshorting of stimulating circuit for tetanic series; M = myograph; its ordinates are verticals, not arcs; T = time; A = contralateral lateral popliteal nerve stimulated, frequency 38 per sec; B, motor nerve stimulated, frequency 38 per sec; T, 0·04 sec (E. G. T. Liddell & C. S. Sherrington (1923), *Proc. Roy. Soc.* B, **95**, 299).

The crossed extension reflex is elicited, as we have just seen, by stimulation of a *contralateral* afferent nerve. If, during the course of such a reflex, an *ipsilateral* afferent is stimulated there is, after a very short latent period, a large reduction of tension (Fig. 45, 11, B). The amount of this inhibition depends on the strength of the stimulation of the ipsilateral afferent nerve but even a weak stimulus can cause some inhibition. Stimulation of the ipsilateral nerve alone, of course, elicits a flexion reflex; this is another way of producing a flexion reflex which in the intact animal is usually elicited by applying a noxious stimulus to the skin. The flexion reflex, no matter how it is elicited, takes precedence over the extension reflex. During the inhibition of a crossed extension reflex, caused by stimulating an ipsilateral afferent nerve, the tension curve falls at the same rate as at the end of a tetanus produced by direct stimulation of the motor nerve. It is probable therefore that the inhibitory stimulus stops the discharge of the motor neurones. This inhibition is quite different from that in smooth muscle which is usually supplied by two separate nerves, stimulation of one producing a contraction and of the other a relaxation. Skeletal muscle has no inhibitory nerve supply and accordingly inhibition in spinal reflexes must be due to events in the spinal cord.

If the stimulation of the contralateral nerve, which produces the crossed extension reflex, is stopped for a short time (Fig. 45, 11, A) the tension in the muscle falls very little because of the after-discharge of the motoneurones. Thus the inhibition seen in Fig. 45, 11, B, cannot be due to the cutting off from the reflex arc of the afferent impulses. The short latent period of the inhibitory effect suggests that it is brought about by impulses which travel from the ipsilateral nerve to the ventral horn cells by a direct route with few synapses.

Sherrington supposes that an afferent impulse which leads to inhibition of a reflex builds up a *central inhibitory state*, or c.i.s., and that this is capable of neutralizing or removing c.e.s. Inspection of the tracings of experiments on inhibition (Figs. 45, 11 and 18) suggests that c.i.s. can last for a considerable time—in some experiments up to 200 msec. The nature of c.i.s. is not yet clear; it may be regarded as a process which makes the polarized membranes more stable perhaps by increasing the potential difference across them, that is hyperpolarization (cf. anelectrotonus, p. 759).

STRETCH REFLEXES

The knee jerk is the most familiar example of this type of reflex, sometimes called a myotatic reflex. It is elicited by tapping the patellar tendon. This stretches the muscle spindles within the quadriceps, the extensor muscle of the thigh, and after a very short latent period the muscle contracts and the leg is jerked forward (Fig. 45, 18). It is essential for the eliciting of the reflex that the quadriceps be initially in a partly stretched condition. The reduced reflex time, about 1·5 msec, is so short that the reflex must be purely spinal and it is probably a monosynaptic reflex. It depends, like any other reflex, on the integrity of the reflex arc which consists of the afferent nerves, spinal cord and the efferent nerves. If the reflex is absent there must be an interruption somewhere in this arc. As the character of this spinal reflex is greatly changed by impulses from higher parts of the brain its modifications are of considerable diagnostic value. They are discussed later in connexion with the motor pathways (p. 894 and also Fig. 47, 10).

In animal experiments the knee jerk is inhibited if a flexion reflex is induced at the same time, as in Fig. 45, 18. Here the period of inhibition outlasts the inhibiting stimulus and may be followed by an increased excitability of the knee jerk as the result of the so-called *rebound* of the motor neurones.

Quantitative studies of stretch reflexes are most easily made by increasing the length of a muscle attached to a myograph. The rise of tension which occurs on stretching is shown extremely well by the extensor muscles of a decerebrate preparation (M, Fig. 45, 19). The rise of tension

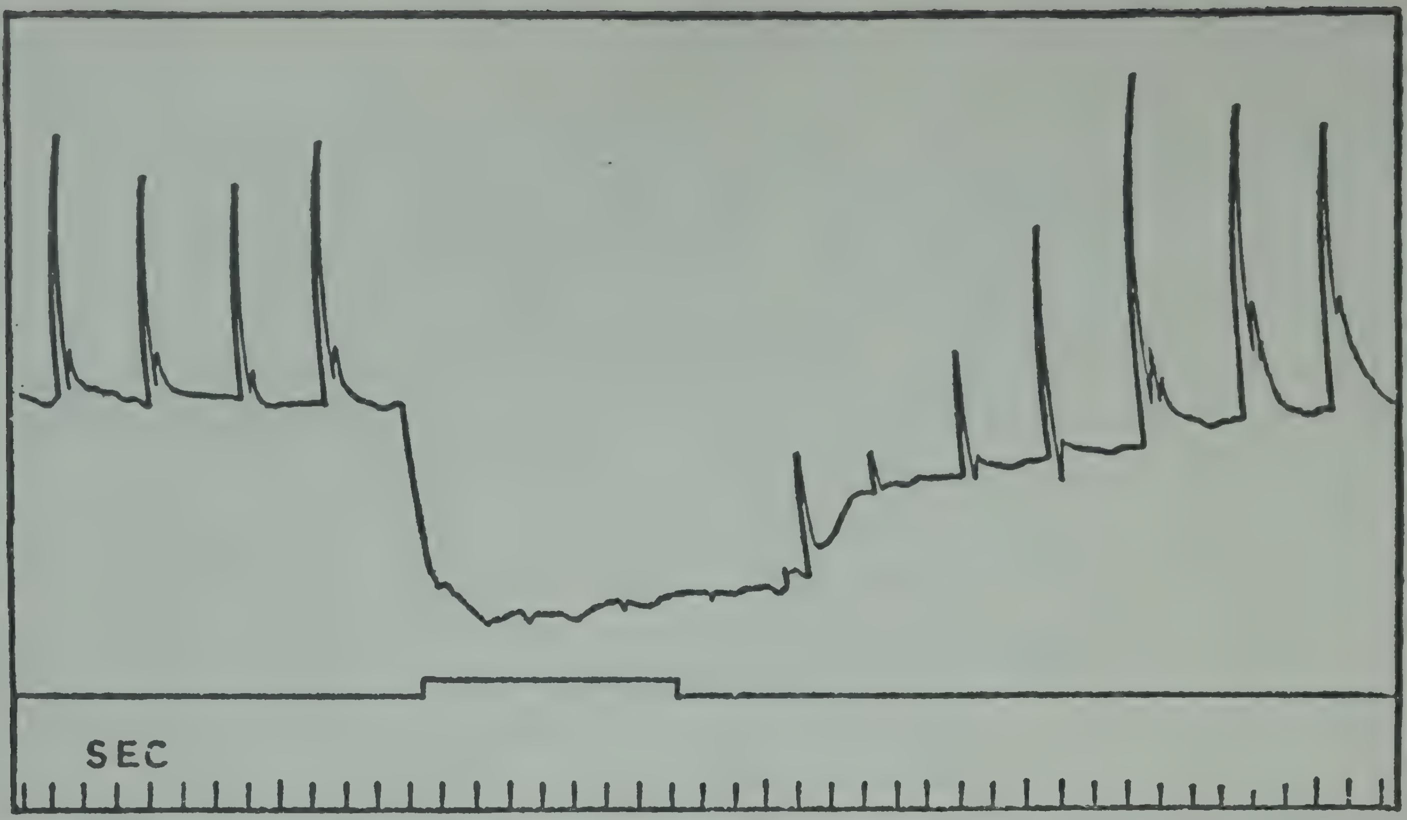


FIG. 45, 18

Reflex inhibition of the knee-jerk. Records from above downwards: Tracing from preparation of the extensor muscles of the knee, recording a series of knee-jerks elicited at each alternate beat of a metronome. Signal line: weak faradization of the central end of the hamstring nerve was applied during the time marked by the rise of the signal. Time in sec. At onset of stimulation the tonus of the extensor muscles at once fell and with it the knee-jerk was temporarily abolished. After cessation of the inhibiting stimulus the tonus and the knee-jerk quickly returned, and the latter became more brisk than before the inhibition. This increase in the reflex was ascribed by Sherrington to a 'rebound' of the motoneurones (C. S. Sherrington (1947), *The Integrative Action of the Nervous System*, p. 89. Cambridge: University Press).

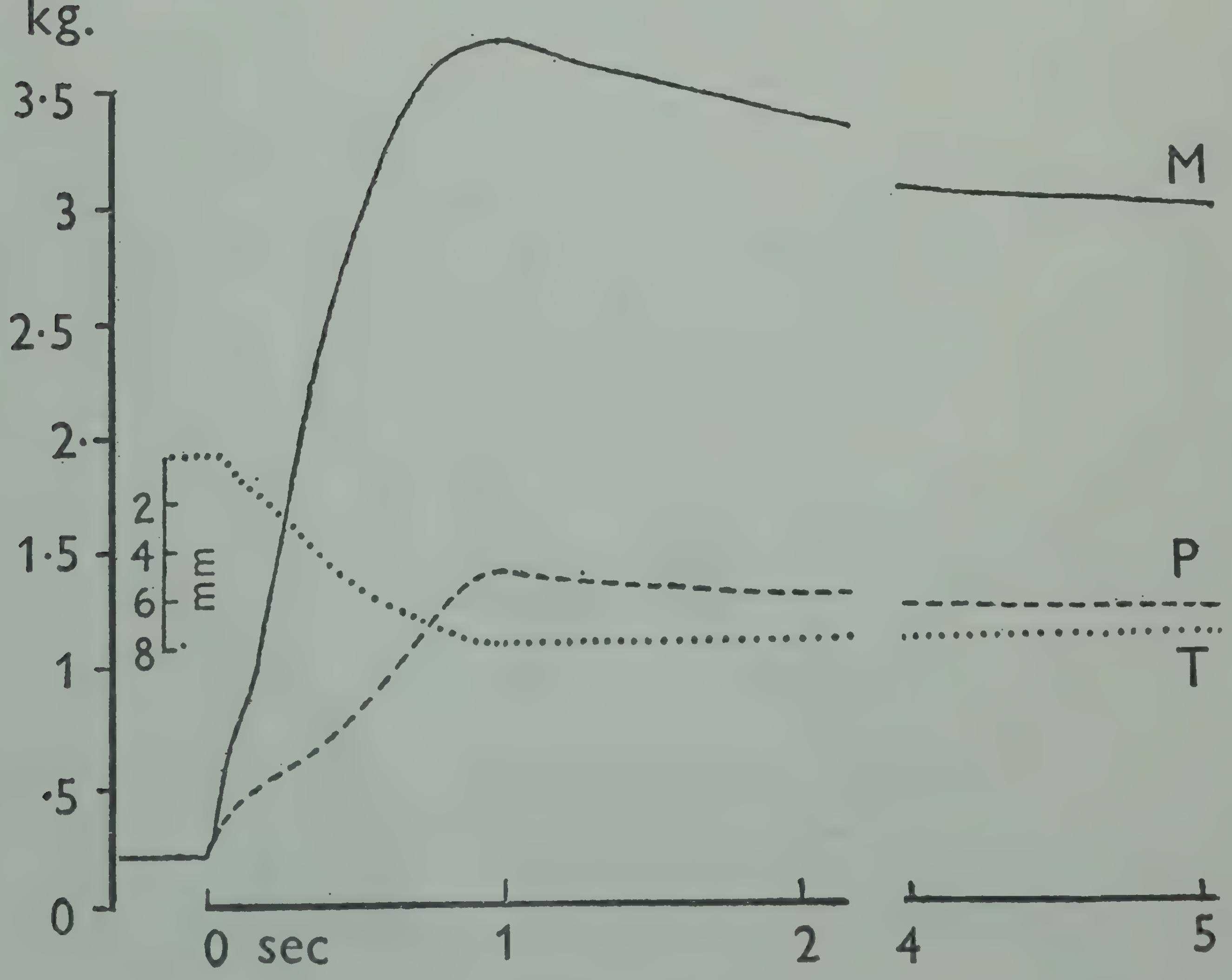


FIG. 45, 19

Whole quadriceps of decerebrate cat. Muscular response to 8 mm stretch T before (M) and after (P) cutting the nerve to the muscle. Myograph multiplies tendon movement 62 times. Time in sec (E. G. T. Liddell & C. S. Sherrington (1924), *Proc. Roy. Soc.* B, **96**, 213).

is dependent on the integrity of the reflex arc but it is not abolished by anæsthetization of the tendon or by removal of the skin. After the nerve to the muscle is cut (P, Fig. 45, 19) the rise of tension on stretching the muscle is relatively small and can be accounted for by the elastic properties of the muscle itself. Since the tension produced by stretching can be reduced by stimulation of an ipsilateral sensory nerve or by stretching an antagonistic flexor muscle there is no doubt about its reflex origin. This myotatic reflex originates by the stretching of receptors, presumably muscle spindles, in the muscle itself. If a quadriceps muscle is divided into four portions which can be separately stretched it is found that only the part extended shows a rise of tension. This can be accounted for by supposing that the impulses from the stretch receptors excite motor neurones which innervate muscle cells in the immediate vicinity of these receptors. In this way each part of the muscle contributes its share to resisting elongation. There is good evidence that the reflex arc for this stretch reflex contains only two neurones, the afferent neurone being of large diameter (see p. 749).

These stretch reflexes are very characteristic of the antigravity extensor muscles and they are especially noticeable in the decerebrate preparation in which the extensor muscles are firmly contracted. That state of contraction of muscles normally present even at rest is described as *tonus* and is reflex in origin, since it is abolished by cutting the anterior or posterior roots or by section of the muscle tendon. A characteristic of muscle tonus and of stretch reflexes is that they can be maintained for long periods with only a small expenditure of energy and with, consequently, an absence of fatigue. It may be that stretch reflexes and postural reactions depend chiefly on red muscle fibres (p. 771).

The stretch reflexes are being intensively studied and much new information about the part played by muscle spindles is available. The muscle spindle itself has already been described on page 771. When a muscle is made to contract by stimulation of the α motor fibres the afferent discharge from the spindles declines, the reflex excitation of the motoneurones to the muscle declines also. When the muscle relaxes at the end of the contraction the spindle discharge increases and the excitability of the α motoneurones is increased. The movement of the muscle is thus smoothed by a reduction of tension during contraction and a lessening of the rate of fall of tension during relaxation. When the muscle is stretched by a pull applied to the tendon the rate of discharge in the large afferent fibres increases and may produce a contraction in the same muscle by a monosynaptic path. The rate of discharge can also be increased, even when the length of the muscle is fixed, by stimulating the γ efferents. These are the small motor nerve fibres which go to the intrafusal fibres (the contractile parts of the spindle) the contraction of which stretches the annulospiral endings. The increased discharge from the latter produces a reflex contraction of muscle until the annulospiral discharge reaches its former value. The discharge in the large afferents, therefore, depends partly on the length of the muscle and partly on the length of the intrafusal fibres.

The function of the γ efferents is not yet clear. By controlling the tension in the muscle spindles they can keep them in a sensitive condition. However, the rate of discharge of the γ efferents is not entirely dependent

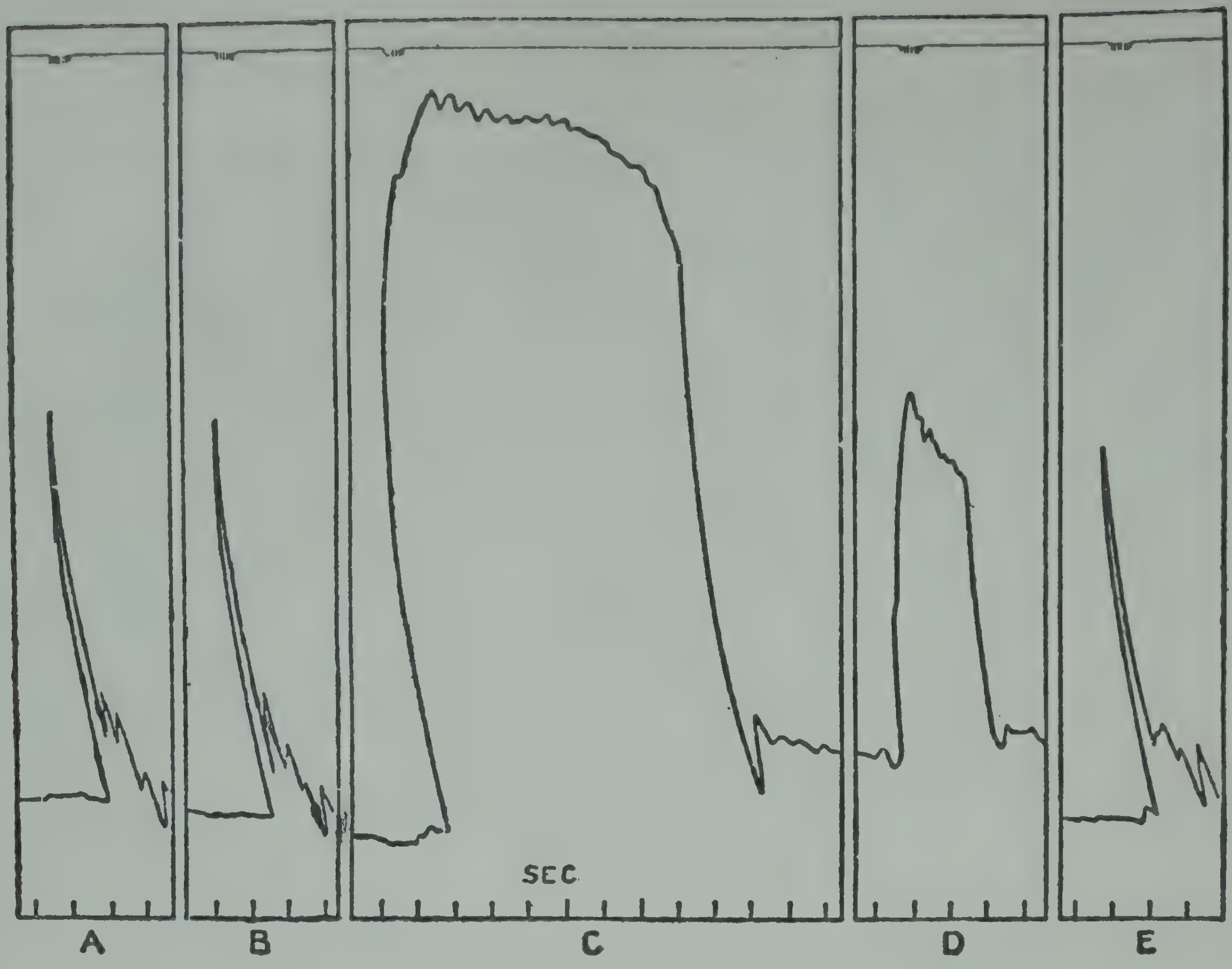


FIG. 45, 20

Successive induction. Crossed extension reflex augmented by precurrent flexion reflex. Records from above downwards: Signal line: break-shocks of each stimulation evoking crossed extensor reflex. Myograph curve: crossed extensor reflex. This reflex was being elicited regularly by eleven break-shocks (unipolar faradization to skin of opposite foot) at 1 min intervals, stimulus and reflex being of low intensity. Time in sec. Intervals of 1 min between A, B, C, D and E. In the interval between B and C a strong flexion reflex of the limb responding in the crossed extension reflex was provoked and maintained for 45 sec. The next following extension reflex C shows augmentation; this augmentation is also evident, though less, in the next crossed reflex D. In E, a minute later, the augmentation is seen to have passed off (C. S. Sherrington (1947), *The Integrative Action of the Nervous System*, p. 210. Cambridge University Press).

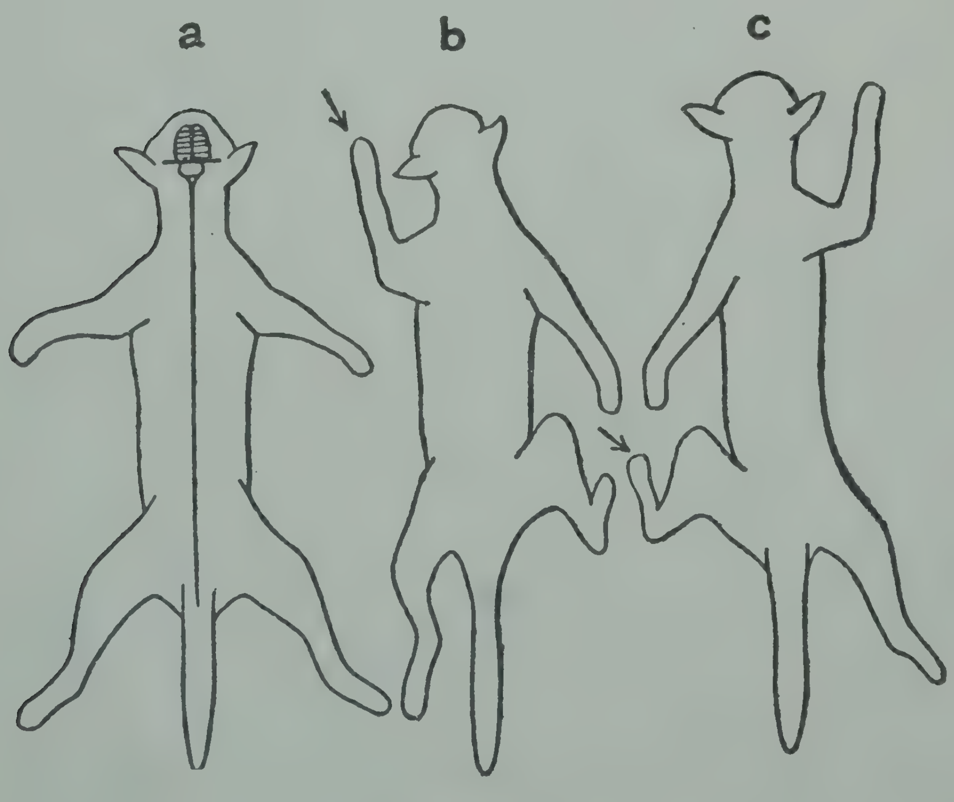


FIG. 45, 21

a = Position under decerebrate rigidity (all four limbs extended). b = Change of attitude from *a* evoked by stimulation of left fore-foot. c = Change of attitude from *a* evoked by stimulation of left hind-foot (C. S. Sherrington (1897–98), *J. Physiol.* **22**, 330).

on proprioceptive reflexes; the rate can be altered by the 'higher centres' and may rise before the α motoneurones discharge. This latter finding has suggested to some investigators that skeletal movements, or at least changes in tone, could be brought about by an increase in the γ efferent discharge, the subsequent increase in the spindle afferent discharge causing through monosynaptic pathways a reflex discharge of the α motoneurones. This pathway seems, however, hardly suited for carrying out rapid movements because of the slow conduction rate in the small fibres and because of the delay in the reflex circuit. Experiments on animals with deafferented limbs or deafferented respiratory muscles have indicated that voluntary and respiratory α motoneurones can be reached directly by descending pathways; the large calibre of the axons of the α motoneurones would be suitable for rapidly executed movements.

A decerebrate preparation (p. 852) with rigidity of the extensor muscles can be used to demonstrate other characteristics of the myotatic reflex. If the experimenter tries to flex the hind-limb at the knee there is first considerable resistance, since the myotatic reflex is evoked, but if the force is maintained the muscle slackens quite suddenly and then can easily be flexed. The sensation is similar to that obtained on closing a pocket-knife and is called by clinicians the *clasp-knife reaction.* Since it involves lengthening of the extensor muscles it may also be called the *lengthening reaction.* This reaction occurs only when great force is applied to the extensor muscle; it may therefore be regarded as a protective mechanism since the muscle gives way before the force is great enough to produce damage by tearing the fibres. The discharge from the Golgi tendon organs, which have a high threshold, is believed to be the basis of the lengthening reaction; the impulses from these organs are inhibitory to the motoneurones of the synergic muscles and perhaps excitatory to the antagonits by a polysynaptic pathway. After the limb of a decerebrate preparation is forcibly flexed in this way the flexed position is maintained.

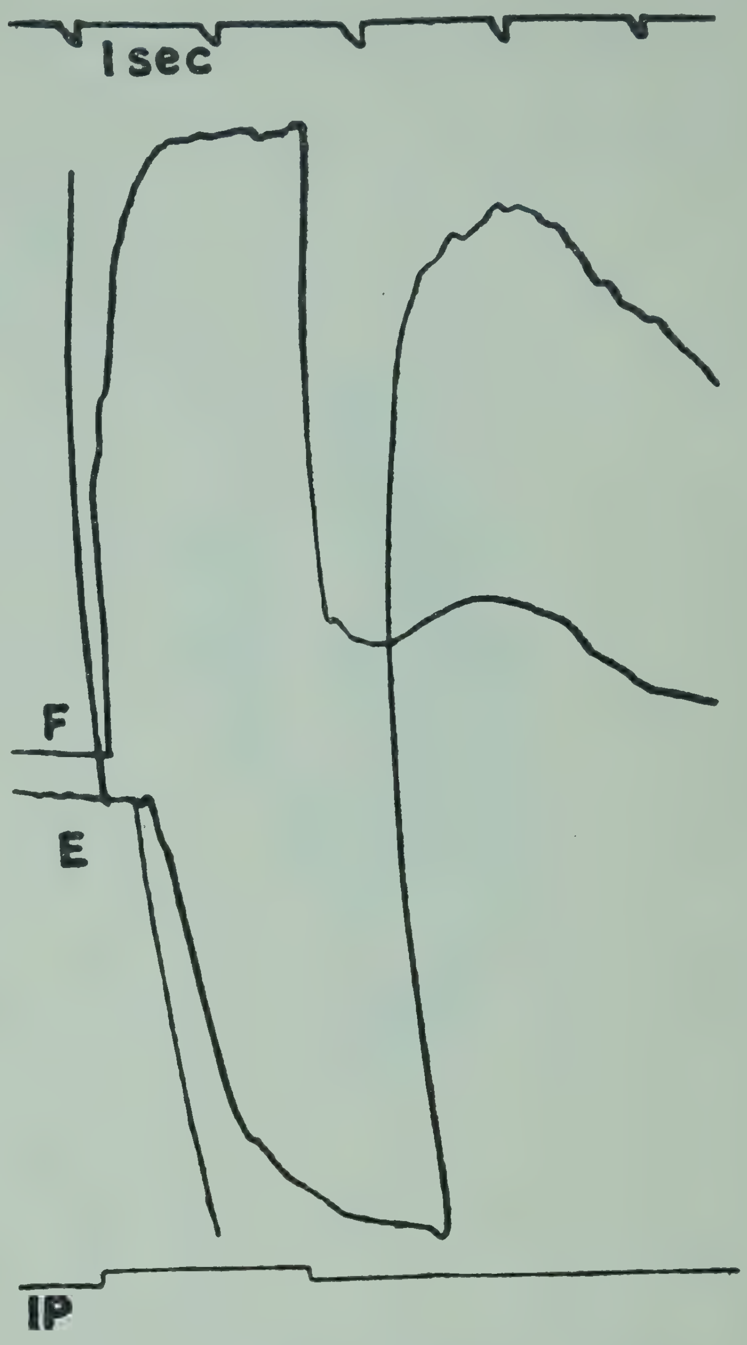


FIG. 45, 22

Reciprocal reflex of antagonistic muscles of knee of a decerebrate cat. F, semitendinosus, a knee-flexor; E, vastocrureus, a knee-extensor. IP, ipsilateral popliteal nerve stimulated during rise of signal line. The flexor muscle contracts and the extensor relaxes. The inhibition of the extensor is followed by marked rebound contraction, which is accompanied by a simultaneous relaxation of the flexor muscle. The myograph writer for extensor muscle is set a little to right of that for flexor muscle, in order that the two may clear each other: the ascent of F and the descent of E are therefore, in fact, practically synchronous (C. S. Sherrington (1913), *Quart. J. exp. Physiol.* **6**, 257).

If the flexed limb is now extended the tension in the extensor muscle is reduced because the ends are brought nearer, and presumably because the γ efferents take up the slack of the intrafusal fibres the afferent spindle discharge increases; the α motor neurone discharge is then increased reflexly so that the position of the limb is maintained. This is called the *shortening reaction*, since the fibres of the extensor muscle are shortened. These two modifications of the stretch reflex give the limbs plasticity, that is to say they allow the limbs to remain in the position in which they are placed. Both reactions are dependent on the integrity of reflex arcs.

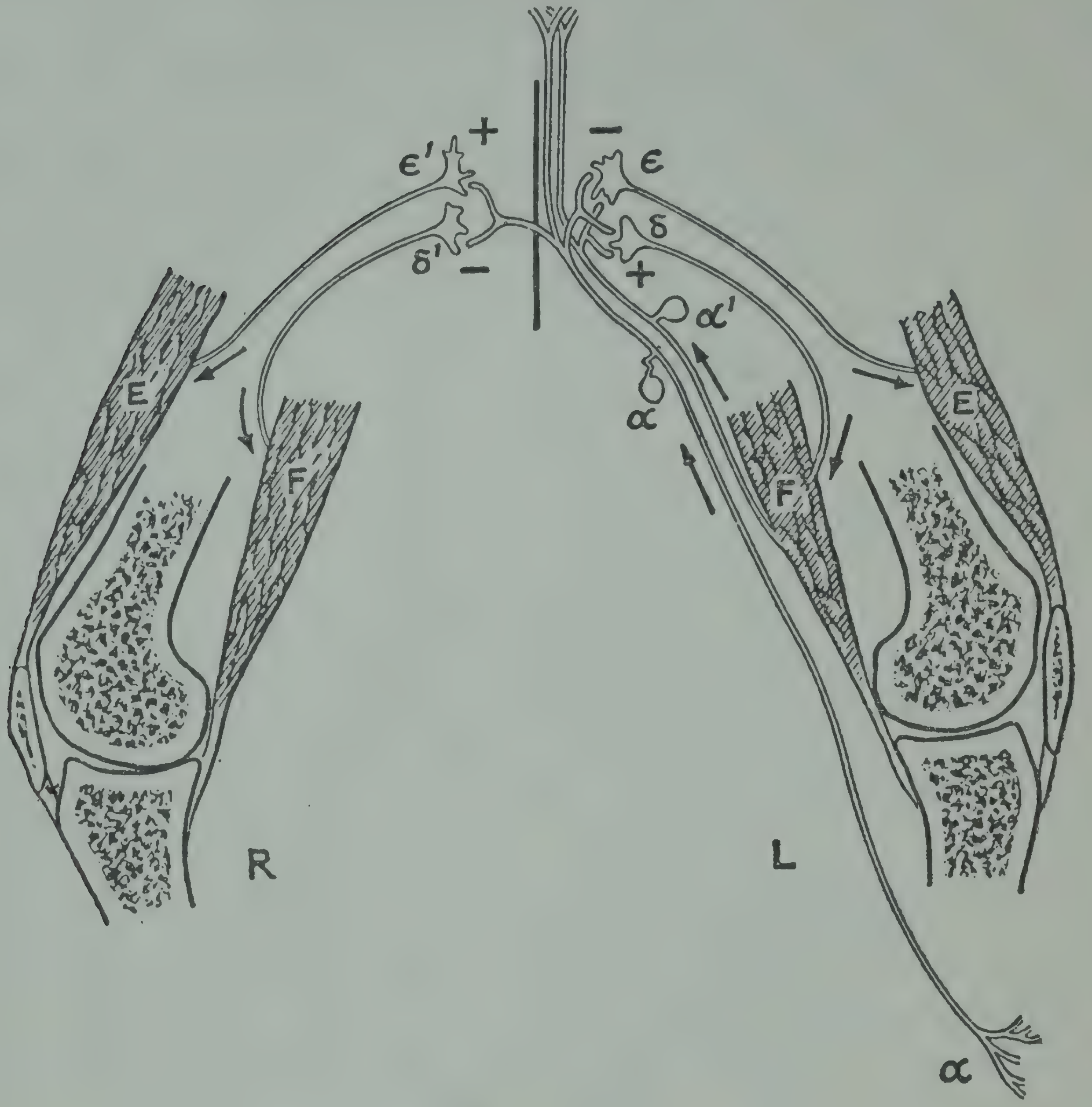


FIG. 45, 23

Diagram indicating connexions and actions of two afferent spinal root-cells, α and α', in regard to their reflex influence on the extensor and flexor muscles of the two knees. α, Cell body of afferent from skin below knee; α', cell body of afferent from flexor muscle of knee, that is, in hamstring nerve; δ and δ', efferent neurones to the flexor muscles of the knee, left and right; ϵ and ϵ', efferent neurones to the extensor muscles; E and E', extensor muscles; F and F', flexor muscles. The interneurones between the afferent and efferent neurones are omitted for simplicity. The sign + indicates that at the synapse which it marks the afferent fibre α (and α') excites the motor neurone to discharging activity, whereas the sign − indicates that at the synapse which it marks the afferent fibre α (and α') inhibits the discharging activity of the motor neurones. The effect of strychnine and of tetanus toxin is to convert the *minus* sign into a *plus* sign (C. S. Sherrington (1905), *Proc. Roy. Soc.* B, **76**, 286).

RECIPROCAL INNERVATION

In addition to these 'static' reflexes, which are responsible for maintenance of posture, the spinal cord is concerned in the 'dynamic' reflexes involved in walking and running. These alternating contractions of the flexors and extensors of the limbs depend on production and inhibition of the appropriate reflexes. At any one instant the extensor muscle group of a limb is contracting and the antagonistic flexor group is relaxing, or vice versa. This involves *reciprocal innervation*, illustrated in Fig. 45, 22, which shows that in a reflex flexion movement there is an active contraction of the flexor muscles with a simultaneous relaxation of the extensor muscles. The converse occurs in a reflex extension movement. Reciprocal innervation is probably due to the division within the spinal cord of each afferent nerve in such a way (Fig. 45, 23) that one branch excites the flexor motor neurones (δ) and a second branch inhibits the extensor motor neurones (ϵ). As the limb flexes the extensor muscles are stretched and there is a reflex contraction of the extensor muscles with a reflex relaxation of the flexors. The resulting extension stretches the flexor muscles in turn and produces flexion of the limb with an inhibition of the extensor muscles. A rhythmic movement is thus set up in which the flexor and extensor reflexes come into action successively and the final common path is commanded alternately. This rhythmic response is readily evoked because after the occurrence of a reflex of one type it is easier to elicit the antagonistic reflex (Fig. 45, 20). This phenomenon is called *negative successive induction*. Fig. 45, 21, illustrates how the alternating movements of the limb which is the well-known pattern of walking may be brought about. At the instant represented in Fig. 45, 21, b, a flexion reflex is elicited in the left fore-limb and an extension in the right fore-limb. This could be explained, as in Fig. 45, 23, by supposing that when an afferent nerve crosses the mid-line its effects on the motoneurones of the flexor and extensor muscles are reversed. By longer connexions in the spinal cord the full walking pattern involving reciprocal activity in all four limbs is obtained.

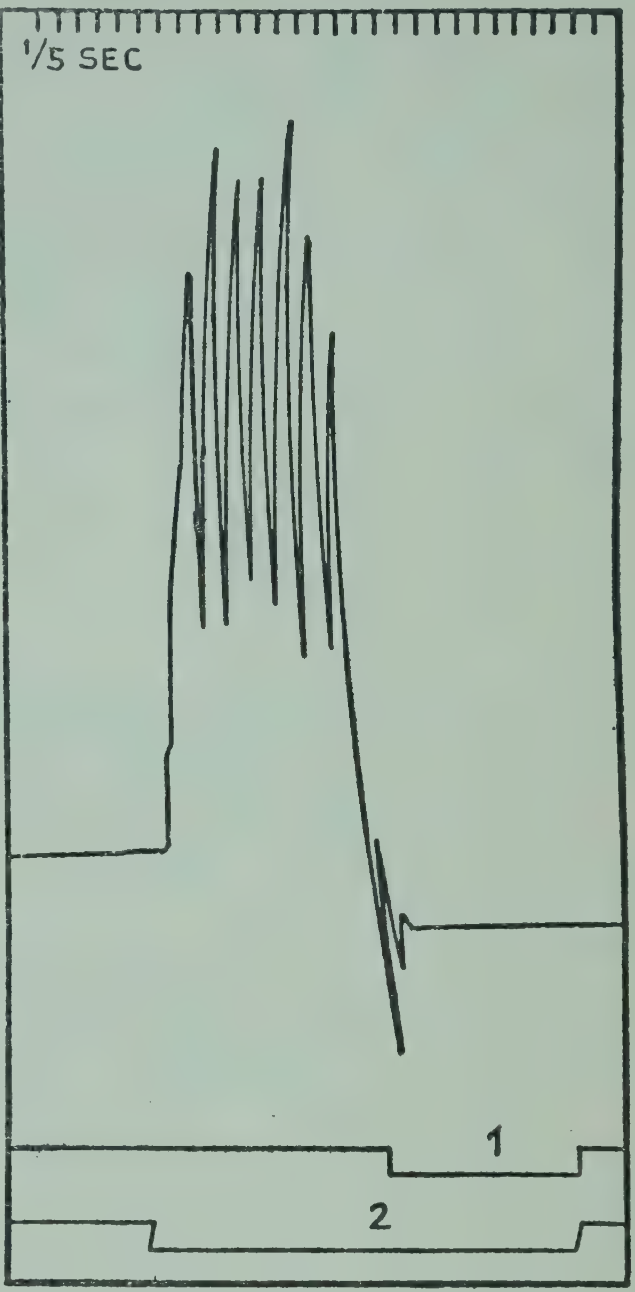


FIG. 45, 24

The scratch reflex cut short by excitation of the skin of a digit of the opposite hind-foot. Records from above downwards: Time in $\frac{1}{5}$ sec. Myograph curve: Scratch reflex. Signal 1: descent of signal marks the period of application of the stimulus to the opposite hind-foot. Signal 2: descent of signal marks the period of application of the stimulus exciting the scratch reflex (C. S. Sherrington (1947), *The Integrative Action of the Nervous System*, p. 193. Cambridge University Press).

The pattern of successive flexions and extensions brought about through reciprocal innervation can be overcome in several ways. For example, when the pad of the foot of an animal is pressed there is a *positive supporting reaction* in which the extensors and flexors contract simultaneously to convert the limb into a pillar to support the weight of the body.

A rhythmic reflex involving reciprocal innervation which has been much studied is the scratch reflex. This can be elicited in a dog which has been allowed to recover after section of the cervical spinal cord below the level of the phrenic nerves. Stimulation of the skin over the ribs results in rhythmic scratching movements (Fig. 45, 24) by the hind-limb which appear purposeful to the observer. At any rate, the movements are accurately enough co-ordinated to remove the source of irritation. Thus this reflex may be said to possess 'localization.' If a noxious stimulus is applied to the foot during a scratch reflex the latter is completely inhibited (Fig. 45, 24). As described already, flexion withdrawal reflexes take the animal away from danger and are prepotent, that is they have precedence over all other reflexes.

REFERENCES

Creed, R. S., Denny-Brown, D., Eccles, J. C., Liddell, E. G. T. & Sherrington, C. S. (1932). *Reflex Activity of the Spinal Cord.* Oxford: Clarendon Press.

Eldred, E., Granit, R. & Merton, P. A. (1953). Supraspinal control of the muscle spindle and its significance. *J. Physiol.* **122,** 498–523.

Fulton, J. F. (1926). *Muscular Contraction and the Reflex Control of Movement.* Baltimore: Williams & Wilkins.

Hunt, C. C. (1952). The effect of stretch receptors from muscle on the discharge of the motoneurones. *J. Physiol.* **117,** 359–379.

Kuhn, R. A. (1950). Functional capacity of the isolated human spinal cord. *Brain,* **73,** 1–51.

Sherrington, C. S. (1947). *The Integrative Action of the Nervous System.* Cambridge University Press.

CHAPTER 46

THE BRAIN STEM AND POSTURAL REFLEXES

THE spinal cord is continuous within the skull with the brain stem which includes the medulla oblongata, the pons and the mid-brain. The medulla includes many important 'centres' such as the respiratory (p. 553), the cardiac (p. 457), and the vasomotor centre (p. 507), and it has also considerable influence on the activity of the digestive tract.

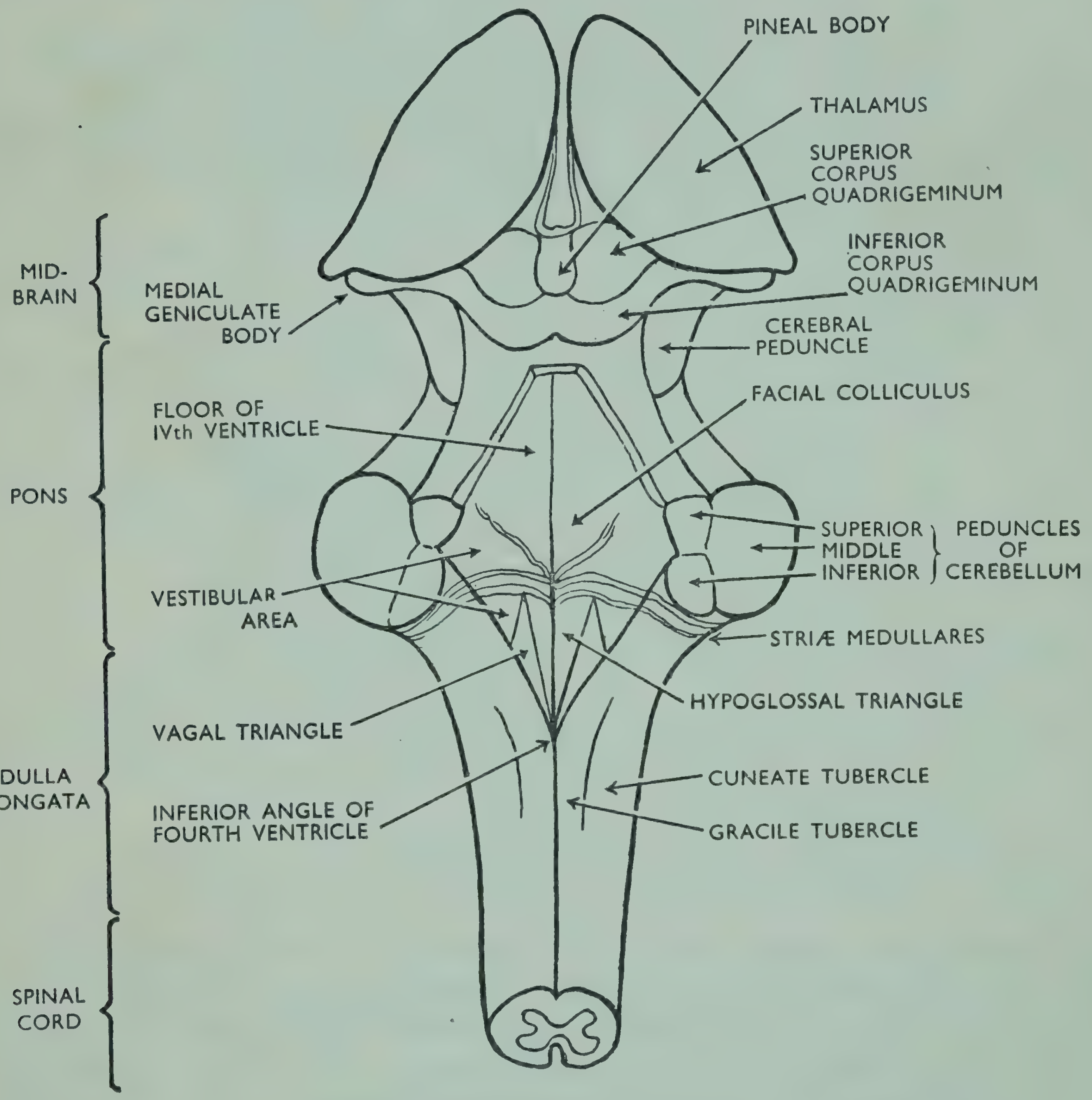

PINEAL BODY
THALAMUS
SUPERIOR CORPUS QUADRIGEMINUM
INFERIOR CORPUS QUADRIGEMINUM
CEREBRAL PEDUNCLE
FACIAL COLLICULUS
SUPERIOR
MIDDLE
INFERIOR
PEDUNCLES OF CEREBELLUM
STRIÆ MEDULLARES
HYPOGLOSSAL TRIANGLE
CUNEATE TUBERCLE
GRACILE TUBERCLE
MID-BRAIN
MEDIAL GENICULATE BODY
FLOOR OF IVth VENTRICLE
PONS
VESTIBULAR AREA
VAGAL TRIANGLE
MEDULLA OBLONGATA
INFERIOR ANGLE OF FOURTH VENTRICLE
SPINAL CORD

FIG. 46, 1

Diagrammatic view of the human brain stem exposed from behind and above (approximately natural size). The cerebellum has been removed by section through the peduncles.

The general arrangement of the spinal cord with its central canal surrounded by grey matter is also seen in the brain stem (Figs. 46, 2, and 46, 3). The plan is, however, complicated by the large nuclei of the cranial nerves, by several crossings (decussations) of fibres and by the widening out and displacement dorsad of the central canal. The medulla oblongata widens as it passes forward (craniad) and the spinal canal opens out into a lozenge-shaped space, the fourth ventricle. On the ventral aspect of the medulla (Fig. 46, 4) the pyramids and their decussation are prominent; a little craniad to this the pons forms a bridge of decussating transverse fibres gathered together into the two middle cerebellar peduncles which consist of fibres running into the cerebellum. On the dorsal aspect of the medulla (Fig. 46, 1) the inferior cerebellar peduncles diverge and pass to the cerebellum. Craniad to the pons is the short mid-brain, the four corpora quadrigemina lying on its dorsal surface. The mid-brain is joined to the forebrain by the two cerebral peduncles.

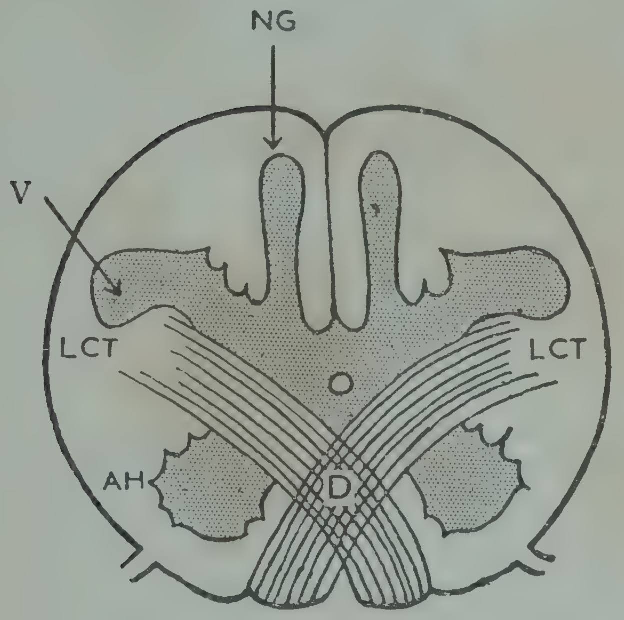


Fig. 46, 2

Section through lowest part of the human medulla at the level of the *Motor decussation.*

AH = most rostral part of the anterior horn.
D = decussation of cerebrospinal fibres (decussation of the pyramids).
LCT = lateral cerebrospinal tract.
V = spinal nucleus of fifth nerve, continuous caudad with the posterior horn of the spinal cord.

(About 3 times actual size.)

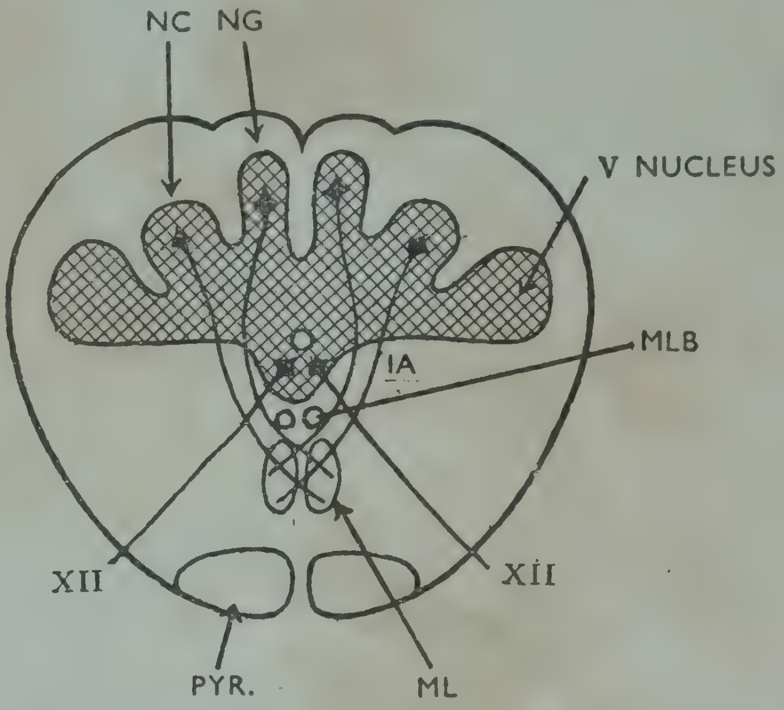


Fig. 46, 3

A section somewhat craniad to that in Fig. 46, 2 to show the *Sensory decussation.*

NG = nucleus gracilis.
NC = nucleus cuneatus.
ML = medial lemniscus.
XII = hypoglossal nerve.
IA = internal arcuate fibres.
PYR = pyramid.
MLB = medial longitudinal bundle (fasciculus).

(About 3 times actual size.)

A transverse section through the caudal, or lower, part of the medulla (Fig. 46, 2) shows that the structure of the spinal cord is still recognizable at this level, but the grey matter is cut across by the fibres of the cerebrospinal tract passing from the pyramids across the mid-line to run down the cord in the lateral corticospinal or crossed pyramidal tract. These interlacing fibres passing obliquely across the cord form the decussation of the pyramids or the *motor decussation.* A section a little craniad to this (Fig. 46, 3) shows the gracile and cuneate nuclei in which the fibres of the posterior columns of the white matter of the cord terminate. From these nuclei new fibres arise and, passing across the mid-line as the internal arcuate fibres, form the medial lemniscus (or medial fillet) which runs to the thalamus. This crossing can be called in physiological terms the *sensory decussation.* The spinothalamic fibres continue to occupy a ventrolateral position, as in the cord, but as they travel craniad, first the touch fibres in the medulla, and then the pain and temperature fibres in

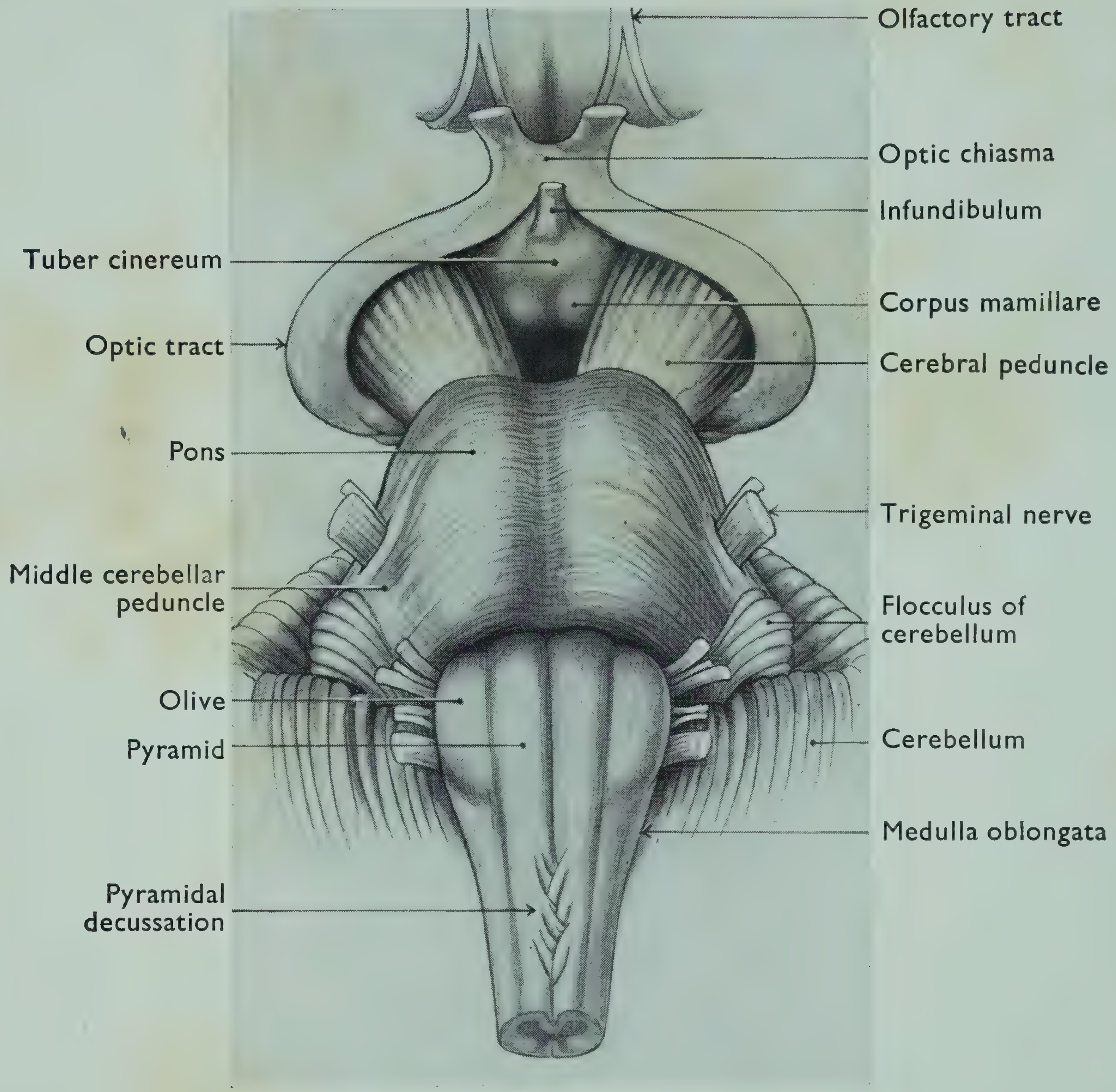

Olfactory tract
Optic chiasma
Infundibulum
Tuber cinereum
Corpus mamillare
Optic tract
Cerebral peduncle
Pons
Trigeminal nerve
Middle cerebellar peduncle
Flocculus of cerebellum
Olive
Pyramid
Cerebellum
Medulla oblongata
Pyramidal decussation

FIG. 46, 4

View of ventral (lower) surface of medulla oblongata, pons and mid-brain of man (approximately natural size).

[*Facing page* 848

the pons pass medially on each side and join the medial lemniscus. In the upper part of the medulla the central canal is displaced towards the dorsal surface of the medulla and opens out to give the appearances shown in Fig. 46, 5.

In the pons (Fig. 46, 6) the pyramidal fibres are separated into small bundles by the fibres of the middle peduncle of the cerebellum which, running at right angles to them, pass across the pons to the opposite cerebellar cortex. The nuclei pontis are small collections of grey matter scattered among the transverse fibres. The superior cerebellar peduncles (Fig. 46, 1) are composed chiefly of fibres coming from the dentate nuclei of the cerebellum which cross in the mid-line (Fig. 46, 7) before entering the thalamus or red nucleus.

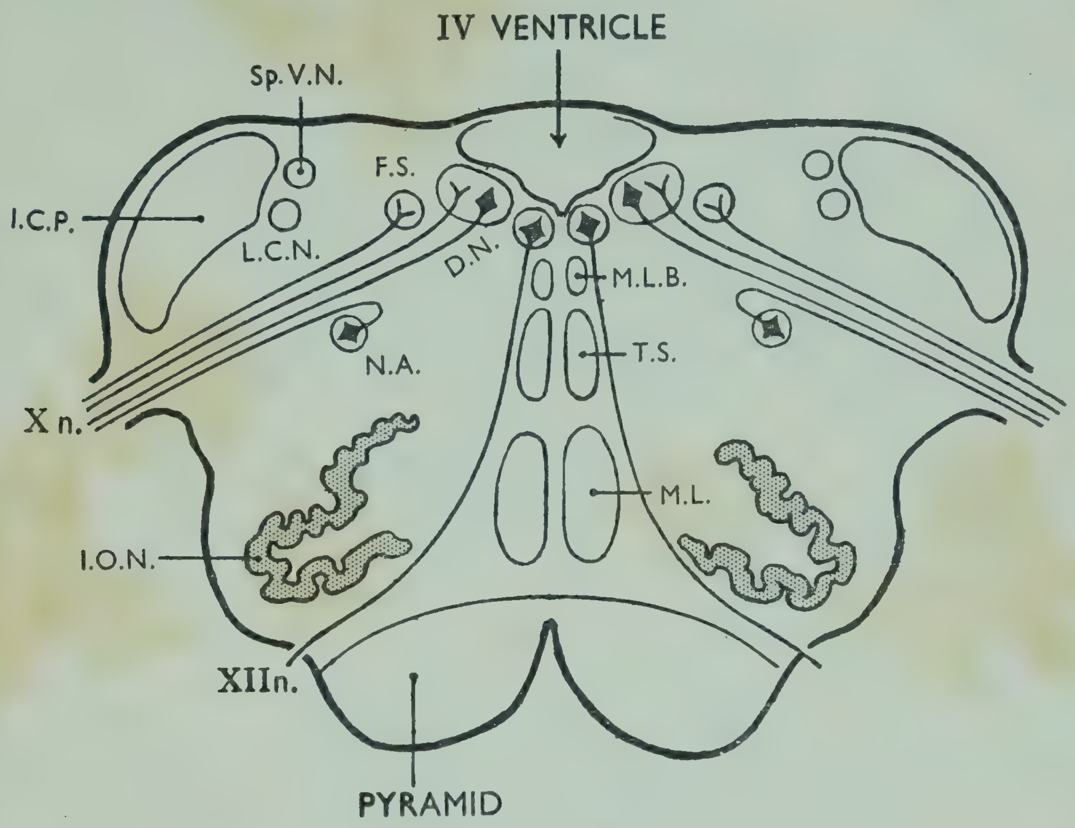


FIG. 46, 5

Section through the upper part of the human medulla oblongata just craniad to the inferior angle of the fourth ventricle. The reticular formation occupies the space between the grey matter on the floor of the fourth ventricle and the pyramids.

I.C.P. = inferior cerebellar peduncle.
Sp.V.N. = spinal vestibular nucleus.
L.C.N. = lateral cuneate nucleus.
N.A. = nucleus ambiguus.
D.N. = dorsal nucleus of vagus.
F.S. = fasciculus solitarius.
M.L.B. = medial longitudinal bundle.
T.S. = tecto-spinal tract.
M.L. = medial lemniscus.
I.O.N. = inferior olivary nucleus.

(About 4 times natural size.)

The mid-brain (Figs. 46, 7, and 46, 8), only about 2 cm long, has a mass of grey matter centrally, around the aqueduct of the mid-brain; in the ventral part of this grey matter lie the motoneurones of the IIIrd and IVth nerves. Dorsal to the aqueduct are the nuclei of the quadrigeminal bodies. A large part of the mid-brain is occupied by the cerebral peduncles which are cut across by the deeply pigmented cells of the substantia nigra into two parts the base ventrally and the tegmentum dorsally. The base carries fibres from the cerebral cortex to the pons and spinal cord. In the tegmentum three decussations occur. The most caudal is the decussation of the superior peduncles of the cerebellum (Fig. 46, 7); the rubrospinal tracts (Fig. 46, 8) and the tectospinal tract cross more rostrally at the level of the IIIrd nerve nucleus. The red nucleus, a prominent feature of the tegmentum in the upper part of the mid-brain, is part of the extrapyramidal system. Fibres from this nucleus cross in the decussation of the rubrospinal tracts and pass down in the formatio reticularis (Fig. 46, 6) of the pons and medulla oblongata to lie in front of the cerebrospinal tract in the cord. The tectospinal and tectobulbar tracts arise in the superior quadrigeminal bodies.

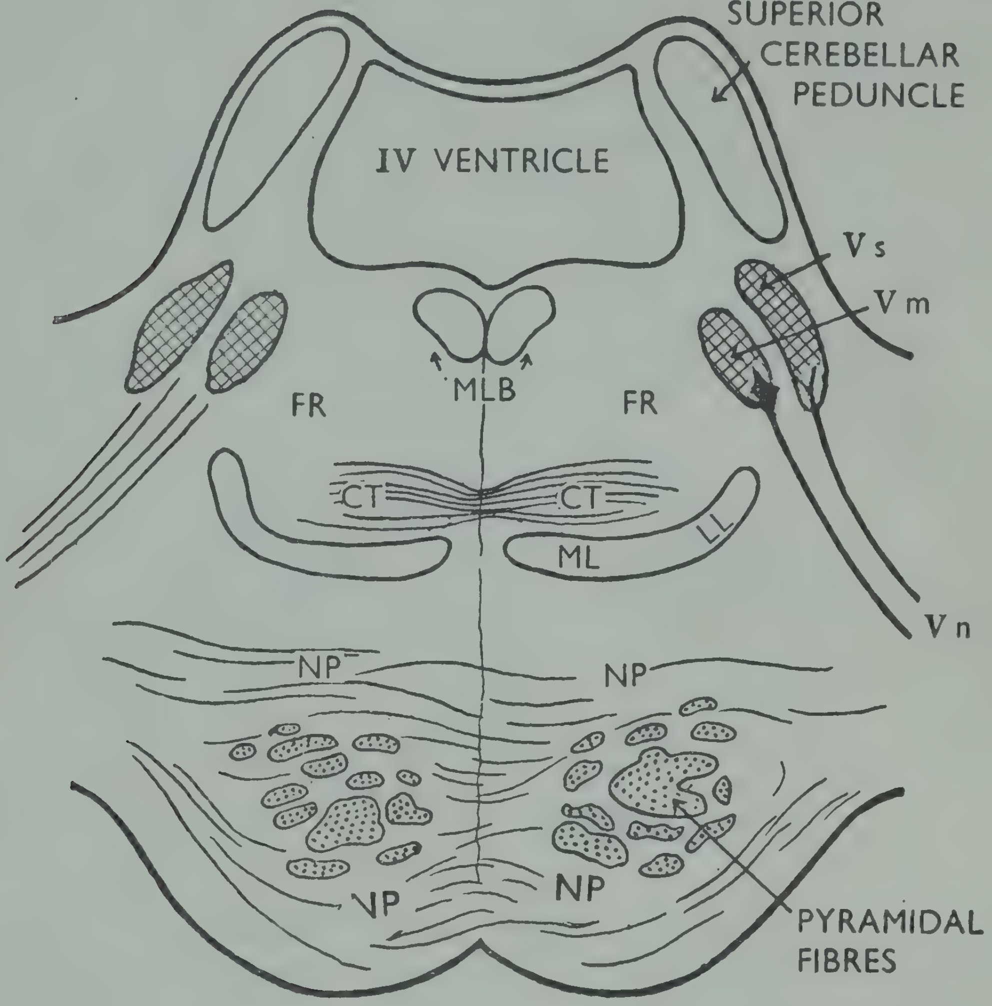


Fig. 46, 6

Section through the upper part of the pons.

FR = formatio reticularis.
NP = nuclei pontis.
ML, LL = medial lemniscus, lateral lemniscus.
MLB = medial longitudinal bundle (fasciculus).
V_s, V_m = sensory and motor nuclei of trigeminal nerve (Vn).
CT = corpus trapezoideum.

(About 4 times natural size.)

Reticular Formation of the Brain Stem

Recent work by Magoun and others has revealed the great importance of the reticular formation of the brain stem (Fig. 46, 6). It consists of scattered grey matter intersected by fibres running in all directions. When the caudal part of the area is stimulated electrically (4 in Fig. 46, 9) in a decerebrate or anæsthetized animal all motor activity ceases whether reflex or cortical in origin. The reticulo-spinal impulses pass down in large fibres in the anterolateral white matter of the cord, chiefly on the same side, to terminate on the spinal interneurones and suppress the motoneurones. An adjacent area (5 in Fig. 46, 9) when stimulated facilitates movements on both sides of the body. The facilitatory fibres pass down the cord quite close to the inhibitory fibres and also act on spinal interneurones. The reticular neurones of the brain stem are played upon by impulses from the

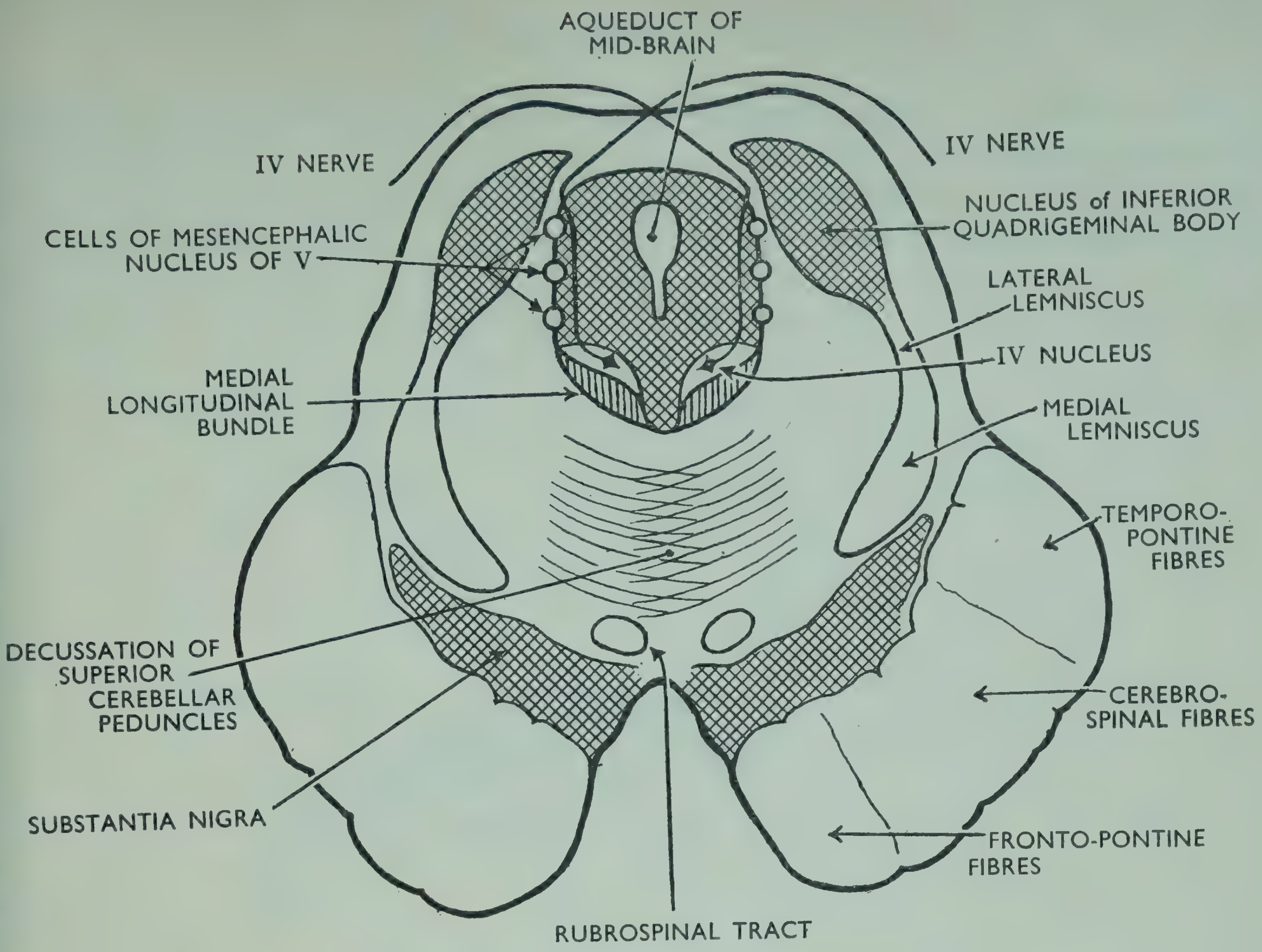

AQUEDUCT OF MID-BRAIN
IV NERVE
IV NERVE
CELLS OF MESENCEPHALIC NUCLEUS OF V
NUCLEUS of INFERIOR QUADRIGEMINAL BODY
LATERAL LEMNISCUS
IV NUCLEUS
MEDIAL LONGITUDINAL BUNDLE
MEDIAL LEMNISCUS
TEMPORO-PONTINE FIBRES
DECUSSATION OF SUPERIOR CEREBELLAR PEDUNCLES
CEREBRO-SPINAL FIBRES
SUBSTANTIA NIGRA
FRONTO-PONTINE FIBRES
RUBROSPINAL TRACT

Fig. 46, 7

Section through the mid-brain at the level of the inferior quadrigeminal bodies. The area between the medial lemnisci is occupied by the reticular formation.

MLB = medial longitudinal bundle (fasciculus).
RS = rubrospinal tract.
SN = substantia nigra.
(About 5 times natural size.)

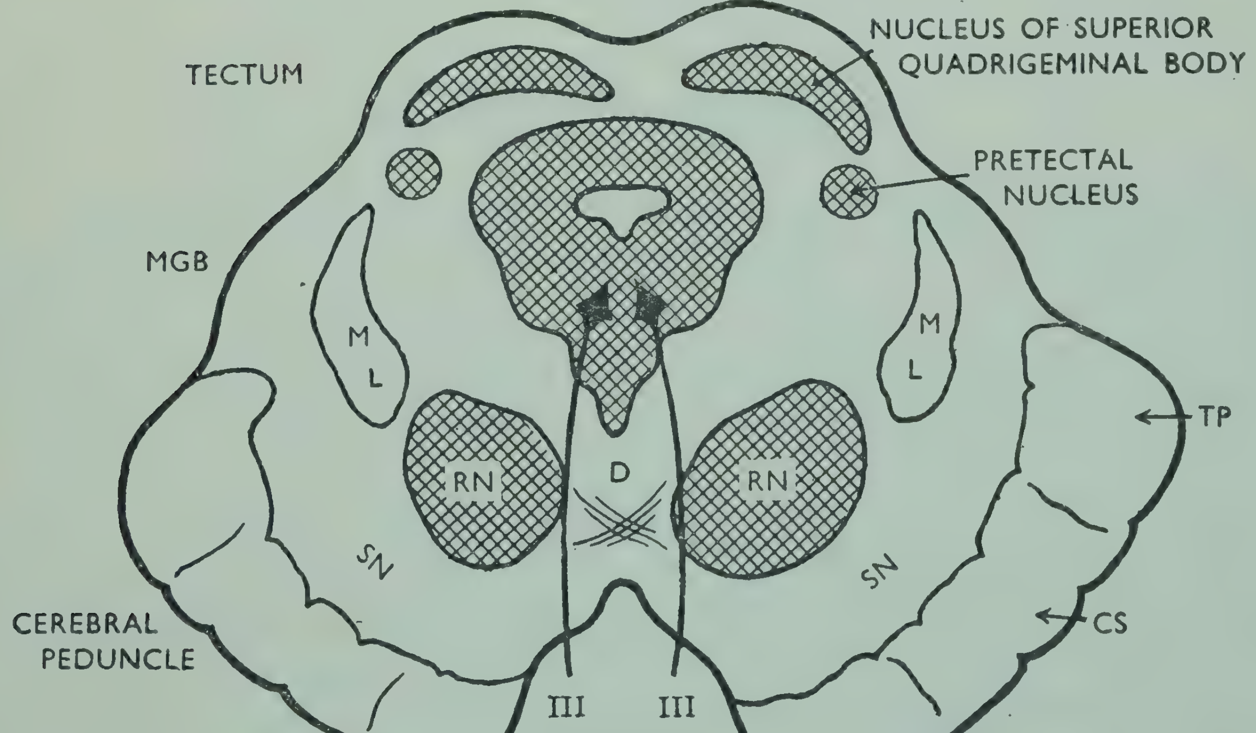

TECTUM
NUCLEUS OF SUPERIOR QUADRIGEMINAL BODY
PRETECTAL NUCLEUS
MGB
M L
M L
TP
RN
D
RN
SN
SN
CEREBRAL PEDUNCLE
CS
III
III
FP

Fig. 46, 8

Section through the mid-brain at the level of the superior quadrigeminal bodies. The pretectal nucleus is actually rostral to the plane of this section. The reticular formation occupies the space between the central grey matter and the substantia nigra.

RN = red nucleus.
MGB = medial geniculate body.
D = decussation of rubrospinal tracts.
ML = medial lemniscus.
SN = substantia nigra.
TP = temporopontine fibres.
CS = cerebrospinal fibres.
FP = frontopontine fibres.
(About 5 times natural size.)

cerebral cortex, caudate nuclei, superior corpora quadrigemina and the cerebellum as shown in Fig. 46, 9. The reticular system also sends impulses rostrally which are concerned in maintaining wakefulness for injury to this area is apt to result in somnolence (p. 908).

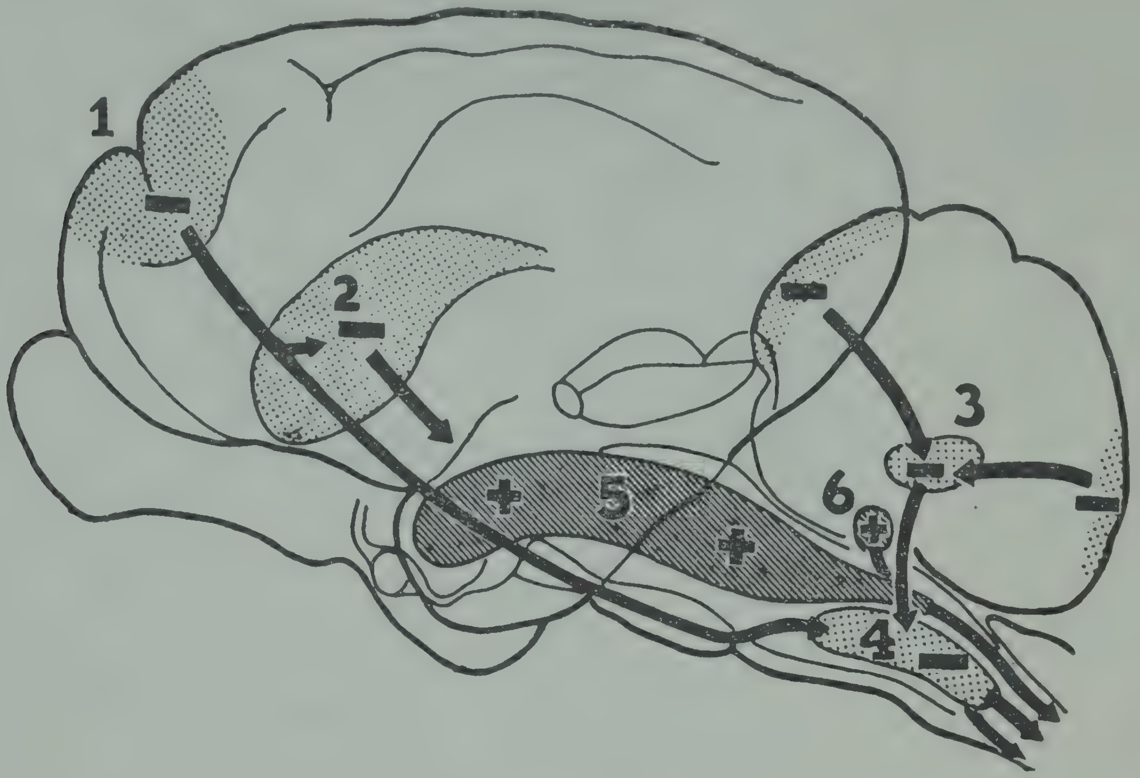


FIG. 46, 9

Reconstruction of cat's brain showing the connexions of the reticular formations of the brain stem. 5, is the facilitatory area and 4, the suppressor area of the brain stem. These areas are played upon by the suppressor area of the frontal cortex 1, the caudate nucleus 2, and cerebellum 3. 6, is the vestibular nucleus which has as described in the text a facilitatory action on spinal reflexes (D. B. Lindsley, L. H. Schreiner & H. W. Magoun (1949), *J. Neurophysiol.* **12**, 197).

DECEREBRATE RIGIDITY

Many of the reflexes described in the previous chapter, for example myotatic reflexes, can be considered as components of the complicated mechanism of posture and locomotion. The modifying and integrating influence of the brain stem on the spinal reflexes has been investigated by making sections through it at various levels to cut off the higher parts of the brain.

If a cut is made between the medulla and the spinal cord the *spinal preparation* shows, after an initial period of flaccidity ('spinal shock,' p. 829), a certain amount of muscular tone not sufficient to bear the weight of the animal; respiration does not occur spontaneously (p. 553) and must be kept up by a pump if the preparation is to survive. A more cranial cut through the medulla produces much the same condition as regards muscle tone but, depending on the exact position of the cut, the animal may be able to breathe spontaneously (p. 553). A section made still further forward through the pons cephalad to the vestibular nucleus, that is between the superior and inferior corpora quadrigemina, produces decerebrate rigidity which was described by Sherrington in the cat in 1898. The limbs of this preparation are extended, the tail is raised and the head elevated. The muscles acting against gravity are rigid so that the animal, if placed in the standing position, is able to maintain this attitude which is, as it were, an exaggeration of the normal posture. The standing position is maintained only if the rigid animal is placed with its four feet on the ground in a stable position. The exaggerated muscle tone is reflex in origin and depends on impulses passing in the dorsal roots, mainly from the muscles themselves. Section of the dorsal roots abolishes the rigidity completely in the hind-limbs but not in the fore-limbs unless the vestibular nerves conveying impulses from the labyrinths are also cut.

The decerebrate animal does not walk, but can be made to walk if supported over a moving platform. It cannot right itself if it is pushed over. Respiration, swallowing and other medullary reflexes are present but the animal cannot maintain its temperature. Such cats have been kept alive for two or three weeks and the rigidity persisted until death. For this rigidity to appear it is necessary for the section to pass caudad to the red nucleus but craniad to the lateral vestibular nucleus which lies under the floor of the fourth ventricle (vestibular area, Fig. 46, 1). A section through the brain stem cephalad to the red nucleus leaves the muscle tone virtually normal. The vestibular nucleus and the vestibulospinal fibres arising from it are essential for decerebrate rigidity, since section of the brain stem caudad to the nucleus or section of the antero-lateral columns of the spinal cord, which carry the vestibulospinal fibres, abolishes it (Fig. 46, 10). Unilateral destruction of the vestibular nucleus abolishes rigidity on the same side (Fig. 46, 11). The cause of the increased rigidity after removal of the cerebellum is discussed below. The cerebral cortex seems not to be deeply involved in the production of decerebrate rigidity since removal of the cerebral hemispheres alone does not produce rigidity unless the animal is lifted off the ground (p. 854). Furthermore, the cerebral cortex controls the muscles on the contralateral side of the body

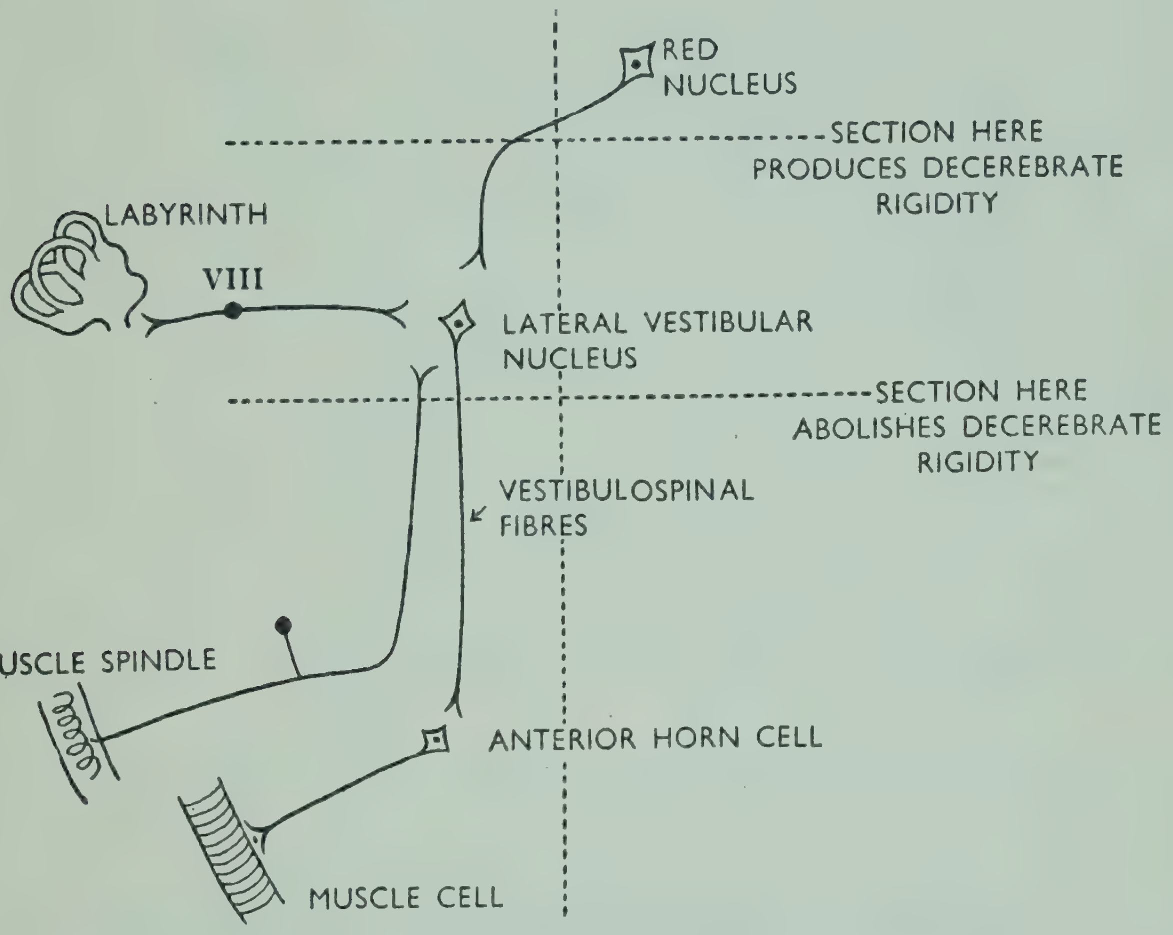

RED NUCLEUS
SECTION HERE PRODUCES DECEREBRATE RIGIDITY
LABYRINTH
VIII
LATERAL VESTIBULAR NUCLEUS
SECTION HERE ABOLISHES DECEREBRATE RIGIDITY
VESTIBULOSPINAL FIBRES
MUSCLE SPINDLE
ANTERIOR HORN CELL
MUSCLE CELL

Fig. 46, 10

Pathways involved in decerebrate rigidity. The section below the red nucleus also cuts through the reticulospinal inhibitory fibres.

whereas unilateral section between the red nucleus and the vestibular nucleus is followed by rigidity on the same side. The rigidity must depend, therefore, on extrapyramidal pathways which go straight down the cord from the pons without any decussation, that is to say on tracts which have already decussated above the level of the section. We are thus led to conclude that decerebrate rigidity occurs in the cat when the vestibular nuclei no longer receive impulses from the higher parts of the brain, which can therefore be presumed to exercise an inhibitory influence on these nuclei. Since destruction of the red nuclei alone does not produce rigidity, the inhibitory impulses cannot be derived from this sourse. It is likely that the source of the inhibitory impulses is the suppressor reticular formation. When this is put out of action by decerebration the activity of the facilitatory reticular system is left unbalanced and decerebrate rigidity occurs. Although a cat deprived of its cerebral cortex (decorticate preparation) has a normal attitude and normal distribution of muscle tone when standing or walking, it shows rigidity when lifted off the ground. Thus the cortex may play a small part in inhibiting the vestibular nucleus although the impulses cannot be conveyed in the pyramidal tracts. Since the decorticate animal does not show rigidity when sitting or standing it is probable that the vestibular nuclei receive inhibitory impulses mainly from parts of the brain other than the cerebral cortex.

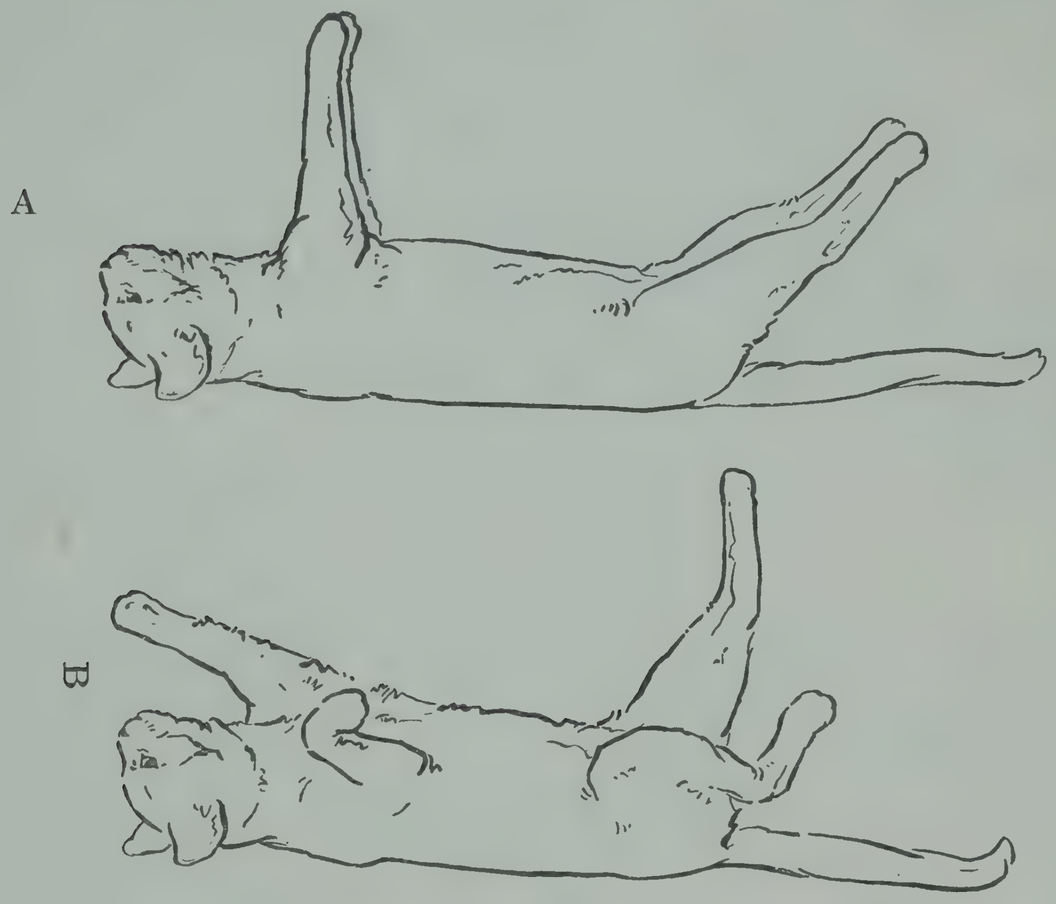


FIG. 46, 11

A, decerebrate cat in the supine position to show rigidity of limbs. B, the same after destruction of the vestibular nucleus on the right side (L. M. N. Bach & H. W. Magoun (1947), *J. Neurophysiol.* **5**, 331).

Recent investigations of decerebrate rigidity show that the γ efferents are specially active. The spindles discharge rapidly and the stretch reflexes are, therefore, very active and in fact are responsible, as already said, for the rigidity. If the dorsal roots of such a preparation are cut, the γ efferents still discharge at the same rate, and the discharge from the muscle spindles is still increased but, since these impulses cannot reach the cord, the rigidity disappears. It was found by Pollock and Davis in 1930, that another type of decerebrate rigidity can be produced by tying off the arterial supply to the brain to make a functional decerebration. In this case deafferentation (by cutting the dorsal roots) does not abolish the rigidity. This second type of rigidity, therefore, could not be brought about by excitation of the γ efferents and must depend on direct activation of the α motoneurones. Later experiments indicate the probable difference between these two types of decerebrate rigidity. An animal decerebrated according to the classical method of Sherrington, which is then deafferented, has flaccid limbs; if now the anterior cerebellum is cooled or cut away rigidity reappears. This can only occur by excitation of the α motoneurones.

Sherrington found that decerebrate rigidity persisted, or even increased, after cerebellectomy. It is now known that spindle activity declines after cerebellectomy and the rigidity must be caused by increased direct excitation of the α pathway, in other words, the operation has caused a switch-over from γ excitation to the α mode of excitation. Indian ink injection of the brains of cats decerebrated by Pollock and Davis's method shows that their procedure corresponds to a section between the forepart of the cerebellum and pons and thus corresponds to the classical decerebrate animal with the anterior cerebellum removed. But the similarity in the reflexes in classical decerebrate animals and in anæmically decerebrate animals suggests that the α and γ systems may often act together; in any case simple observation cannot distinguish which is in operation. The cerebellum may control in some way the kind of activity predominating at any given time.

The pattern of decerebrate rigidity has been described as a caricature of normal standing for several reasons. In normal standing both extensors and flexors are contracted to make the limbs into pillars (the positive supporting reaction, p. 846); in decerebrate rigidity, however, only the extensors are contracted and the contraction is maintained whether the animal is standing on a level surface, on an inclined plane, or lying on its side, that is to say, it is independent of the animal's position. The decerebrate animal is unable to carry out the correcting movements necessary to maintain balance and if it topples over it cannot right itself.

Tonic Neck Reflexes

The decerebrate cat has been used, especially by Magnus and de Kleijn, to study postural reflexes which arise in the neck muscles and labyrinths. To investigate the tonic neck reflexes it is necessary to exclude impulses from the labyrinths by performing a double labyrinthectomy; on the other hand the labyrinthine reflexes can be studied satisfactorily only when the influence of the neck receptors is removed by cutting through the upper cervical dorsal roots or by immobilizing the neck in a plaster of Paris collar.

When the head of a decerebrate cat is turned to the right the right limb is extended and the left limb becomes relaxed. If the head is tilted so that the nose is elevated the fore-limbs extend and the hind-limbs relax, as in Fig. 46, 12, A. The opposite effect is seen when the nose is directed downwards (Fig. 46, 12, B). Pressure with the hand over the cervical vertebræ causes a relaxation of all four limbs (Fig. 46, 12, C). These alterations in the position of the head give rise to afferent impulses which enter the central nervous system by the first three cervical dorsal roots to initiate reflex alterations of tone in distant parts of the body and produce attitudes often observed in the normal dog and cat.

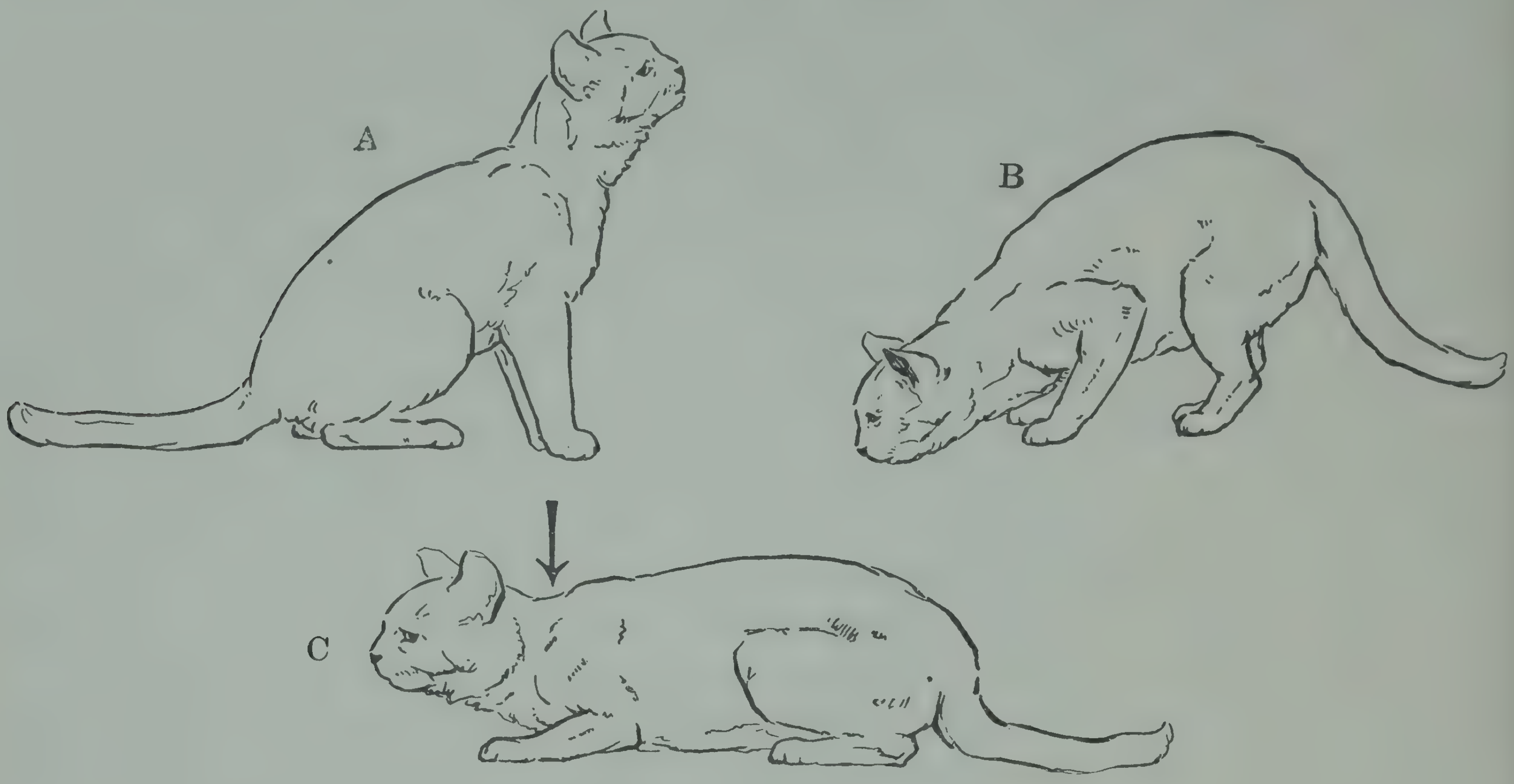


Fig. 46, 12 (A, B and C)

Diagram to show the influence of neck reflexes on posture. Attitude of decerebrate animal when the head is pointed up, A, and pointed down as in B, and when pressure is applied to the neck as at C. Position A is seen in a normal animal when 'begging' at a table; position B when looking for a mouse under a dresser.

THE LABYRINTHS

The bony labyrinth consists of three cavities, the vestibule, the semicircular canals and the cochlea, hollowed out of the petrous part of the temporal bone, and containing clear fluid called perilymph. The membranous labyrinth (Fig. 46, 13), consisting of the duct of the cochlea, the utricle and saccule and the three semicircular ducts, lies within the bony labyrinth and contains fluid called endolymph. The superior and posterior canals lie in vertical planes at right angles to one another while the lateral or horizontal canal is in a plane passing backwards and slightly down; the superior canal of one side is in a plane approximately parallel to the plane containing the posterior canal of the other side (Fig. 46, 18, A and B). The semicircular ducts, each about 15 mm long in man, are smaller in diameter than the bony semicircular canals which enclose them; each duct has a wider portion or ampulla containing the ampullary crests (*cristæ ampullares*), consisting of fusiform supporting cells with hair cells lying between them. A mass of mucus, the cupola, lies on the hair cells which are in close contact with axons of the vestibular nerve. The saccule and utricle each possess a *macula* consisting of supporting cells and hair cells, the ends of which project into the endolymph; rounded bodies composed of calcium carbonate called otoliths lie suspended in the endolymph in contact with the projecting ends of the hair cells. Nerve filaments run from the hair cells to the vestibular nerve and the whole arrangement is

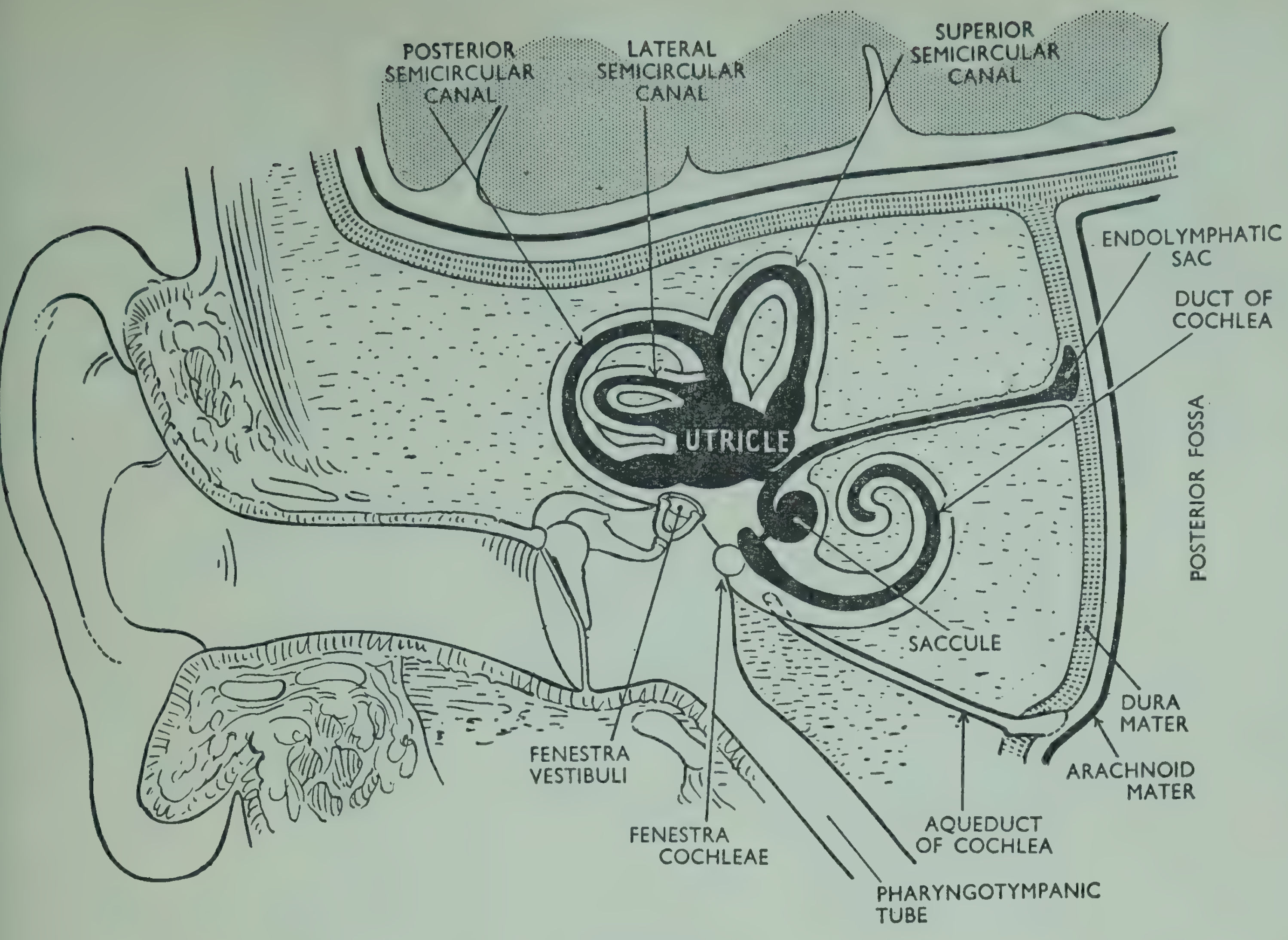

POSTERIOR SEMICIRCULAR CANAL
LATERAL SEMICIRCULAR CANAL
SUPERIOR SEMICIRCULAR CANAL
ENDOLYMPHATIC SAC
DUCT OF COCHLEA
UTRICLE
POSTERIOR FOSSA
SACCULE
DURA MATER
ARACHNOID MATER
FENESTRA VESTIBULI
FENESTRA COCHLEAE
AQUEDUCT OF COCHLEA
PHARYNGOTYMPANIC TUBE

Fig. 46, 13
The bony and membranous labyrinths.

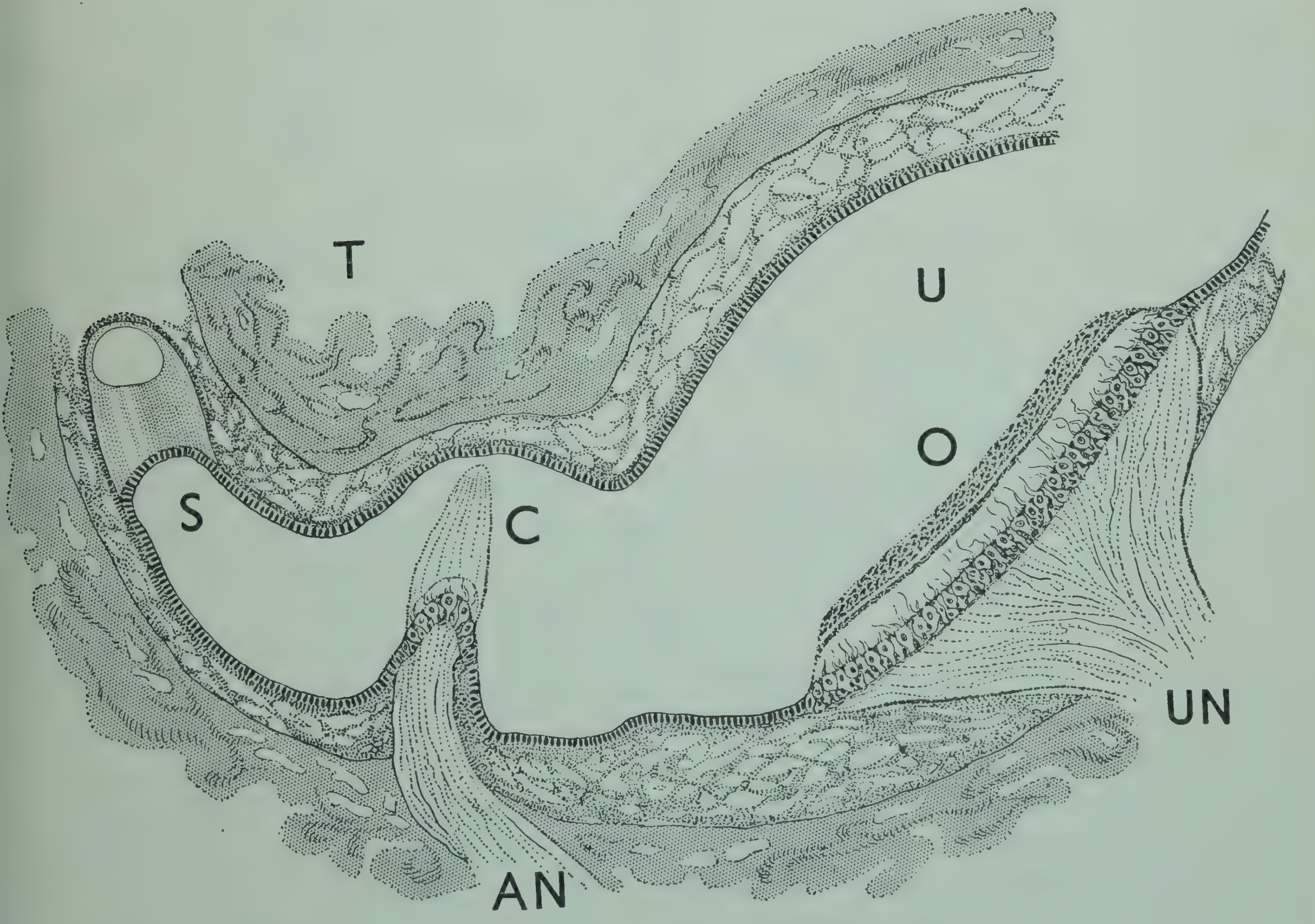


Fig. 46, 14
A diagrammatic cross-section through the utricle U and one ampulla. Both the crista ampullaris C and the otolith organ O are covered with mucous material projecting into the endolymph. The mucus in the case of the otolith organ contains calcareous particles. T, temporal bone. S, semicircular duct. AN, ampullary nerve. UN, utricular nerve. (*After* F. Tandler (1929), *Lehrbuch der systematischen Anatomie, Vol.* 4, *Nervensystem und Sinnesorgane.* Berlin: Vogel.)

known as an *otolith organ* (Fig. 46, 14). In the erect position in man the utricular maculæ are approximately horizontal so that the hair cells project vertically upwards with the otoliths lying on them. The saccular maculæ lie in a vertical plane; it is doubtful if they are functional in mammals.

The fine nerve filaments from the hair cells in the cristæ and maculæ become myelinated as they pass centrally in the vestibular nerve, a division of the eighth cranial nerve, to the vestibular ganglion and then to the vestibular nuclei, as shown in Fig. 46, 15. From there fibres run in the median longitudinal bundle to the oculomotor nuclei through which changes in position of the eyes are effected. Some fibres also pass in the medial lemniscus to the thalamus; others go in the inferior cerebellar peduncle to the cerebellar cortex, chiefly to the flocculonodular lobe.

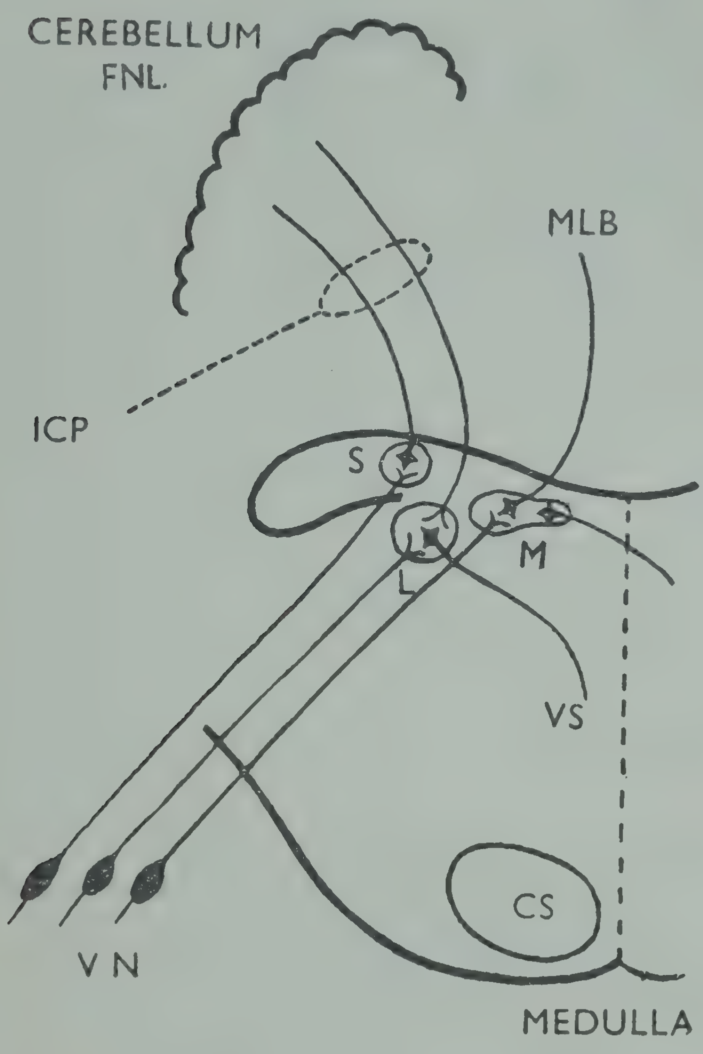


FIG. 46, 15

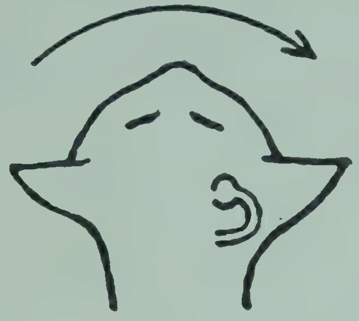
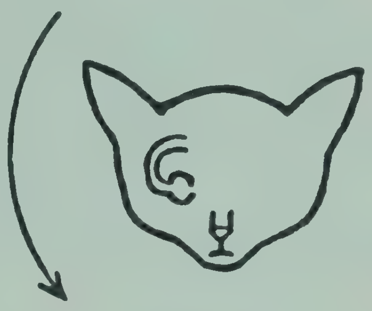
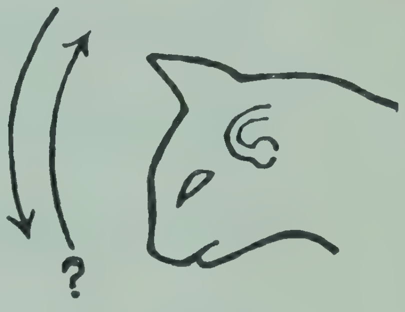
Diagram to show the connexions of the vestibular fibres.

S, L, M = superior, lateral and medial vestibular nuclei.
VS = vestibulospinal fibres.
CS = cerebrospinal fibres.
VN = ganglion of vestibular nerve.
ICP = inferior cerebellar peduncle.
MLB = medial longitudinal bundle (fasciculus).
FNL = flocculonodular lobe.

When the head is rotated the perilymph and endolymph in the semicircular canal lying in the plane at right angles to the axis of rotation tend to lag behind the movement of the head; the cupola is, therefore, deflected and the hair cells are stimulated. In this way the semicircular canals give information about the angular acceleration of the head in space. If the rotational movement is continued to reach a steady angular velocity the lymph moves with the canal, the cupola returns to its normal position and the stimulation of the hair cells ceases. When the rotation of the head ceases the fluid, owing to its momentum, continues to move and deflects the cupola in the opposite direction. The cupola can be thought of as resembling a swing door or a highly damped pendulum. This view of the working of the canals has been confirmed by recent electrophysiological investigations.

Experiments on frogs and fish show that the cristæ of the semicircular ducts signal angular acceleration by an alteration in the resting frequency of discharge, a rotation of the head in one direction increasing the frequency of discharge and a rotation in the other direction reducing it. In the cartilaginous fishes the flow of endolymph and perilymph can be prevented by ligation of the semicircular canal or by cutting through it. These operations leave the resting discharge unaffected but abolish the responses to rotation. From these and other experiments it is fairly clear that it is the deflexion of the cupola which is responsible for the alteration in the resting discharge.

It is possible in certain cold-blooded animals to make records of discharges of single nerve-endings in the maculæ. By peeling away fibres arising in the utricle until only one or two are left Loewenstein and Sand have shown that in the skate the nerve-endings in the macula, when the position of the head is normal, discharge steadily at a rate of 6 to 12 per second at 20° C. When the preparation is tilted the discharge rate either increases or decreases according to the direction of tilt and the new discharge rate continues so long as the new position is maintained. The otolith organs, therefore, signal static position by changing the frequency of discharge according to the gravitational pull on the otoliths with respect to the hair cells; they can thus be thought of as 'out-of-level indicators.' Linear acceleration, like gravity, causes a force to act on the otoliths which, therefore, signal this kind of movement and perhaps also fast angular acceleration. The otolith organ is a rather inert receptor since it responds only to lasting or strong stimuli.

	 A	 B	 C
Rotation in	horizontal plane	transverse plane	median plane
Canal stimulated	horizontal	posterior	anterior
Increase in frequency of impulses when	ampulla trailing.	ampulla leading.	? both leading and trailing (confusion with utricular otolith possible).

In all cases the arrow indicates the direction of rotation producing an increase in the frequency of impulses. Rotation in the other direction produces a fall in the discharge rate.

FIG. 46, 16

Table to illustrate the effects of rotation of the head (in the cat) on the impulses arising in the semicircular canals (E. D. Adrian (1943), *J. Physiol.* **101,** 397).

It has been shown by Adrian that the findings in the lower vertebrates are applicable to the mammalian labyrinth. Using very small electrodes he picked up impulses from the vestibular nucleus in the floor of the fourth ventricle (Fig. 46, 1) about the region of the striæ medullares. This nucleus receives fibres from the cristæ of the ampullæ as well as from the maculæ. The responses obtained are of two different kinds, presumably arising in two different kinds of end-organ. One form of response depends on the position of the head in space and signals tilting and linear acceleration; it consists of slowly adapting trains of impulses which probably arise from gravitational pull on the otoliths. The second form of response, the rotation-controlled type, is produced by angular acceleration and is unaffected by the position of the head. These impulses are presumably derived from the cristæ of the semicircular ducts since they are most sensitive to rotation in one particular plane; rotation in planes at right angles produces little or no alteration of resting discharges. Rotation in one direction increases the frequency of discharge, rotation in the opposite direction diminishes the frequency (Fig. 46, 16). If a rotation which produced an increase in frequency is arrested the discharge rate declines and may fall to zero; it usually takes ten to thirty seconds to regain the normal resting discharge rate and this is presumably the time taken by the cupola to regain its resting position. These findings have been confirmed by Gernandt who inserted a needle electrode into the vestibular (VIIIth) nerve of the cat just as it emerges from the internal auditory meatus (Fig. 46, 17). The discharge from a single nerve of the left horizontal canal, for example, was seen to be increased by rotation to the left, as Adrian had found (Fig. 46, 16). If a record of the activity of the same nerve was made during injection of 20 ml. of cold water into the external meatus the discharge was reduced or stopped; injection of hot water into the meatus increased the frequency of the nerve impulses.

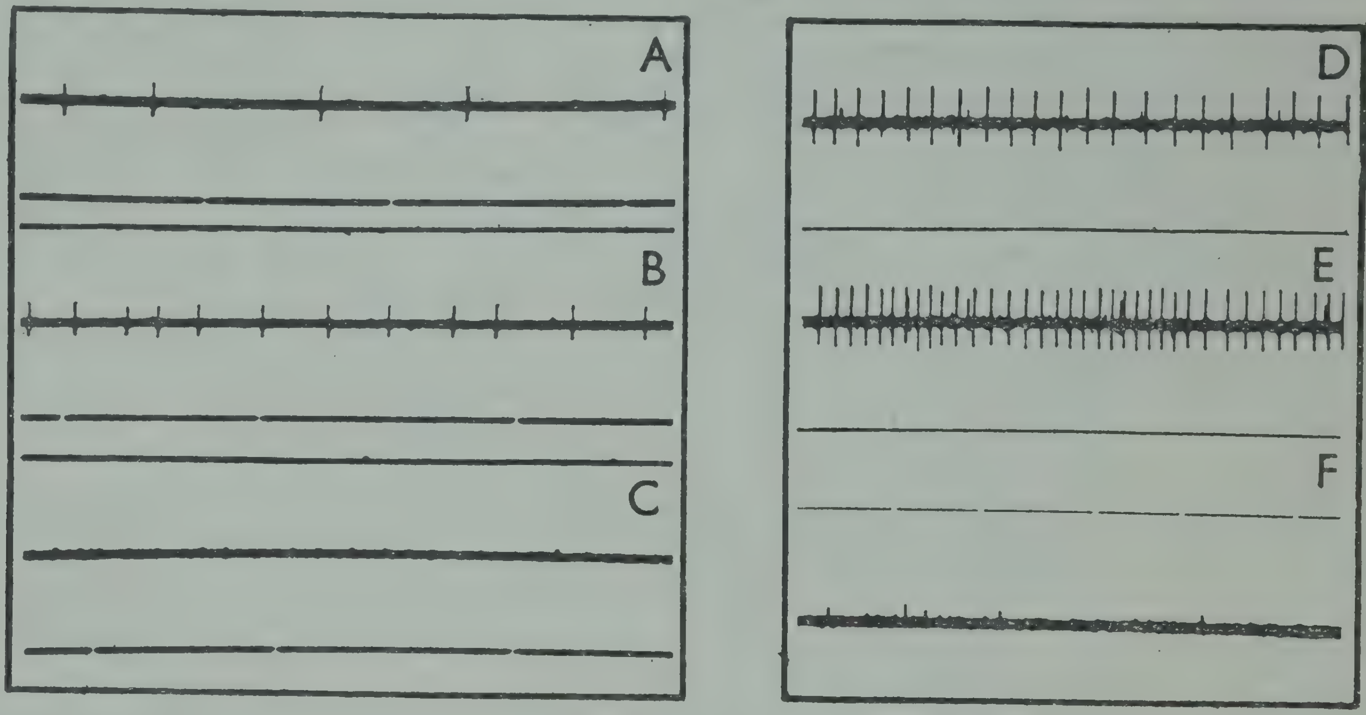


FIG. 46, 17

Decerebrate cat. Discharge of unit activated by horizontal semicircular canal, left side. Semicircular canal horizontal, (A) spontaneous activity; (B) during counter-clockwise rotation; (C) during clockwise rotation; (D) spontaneous activity; (E) caloric stimulation with water at 43° C; (F) same, water at 20° C. Time 0·2 sec. (From R. Gernandt (1949), *J. Neurophysiol.* **12,** 177.)

Bilateral removal of the labyrinths in lower animals produces a very unsteady posture with great loss of muscular tone and ataxia, but in the higher animals this soon disappears as the animal learns to compensate for its disability by using its eyes. The unsteadiness returns if the animal is blindfolded. A similar phenomenon has recently been described in man after treatment with streptomycin which damages the vestibular nuclei and the vestibular apparatus. Unilateral labyrinthectomy in lower animals leads to unequal tone in the neck muscles and the neck righting reflexes are brought into continuous action so that bending movements and rotation of the trunk and jerking movements of the eyes (nystagmus) occur. These disappear very soon. In man there is only a transient disturbance of posture after section of the vestibular nerve because in man, as in the higher apes, muscle-joint sense and vision play a more important part in posture than do the labyrinthine proprioceptors.

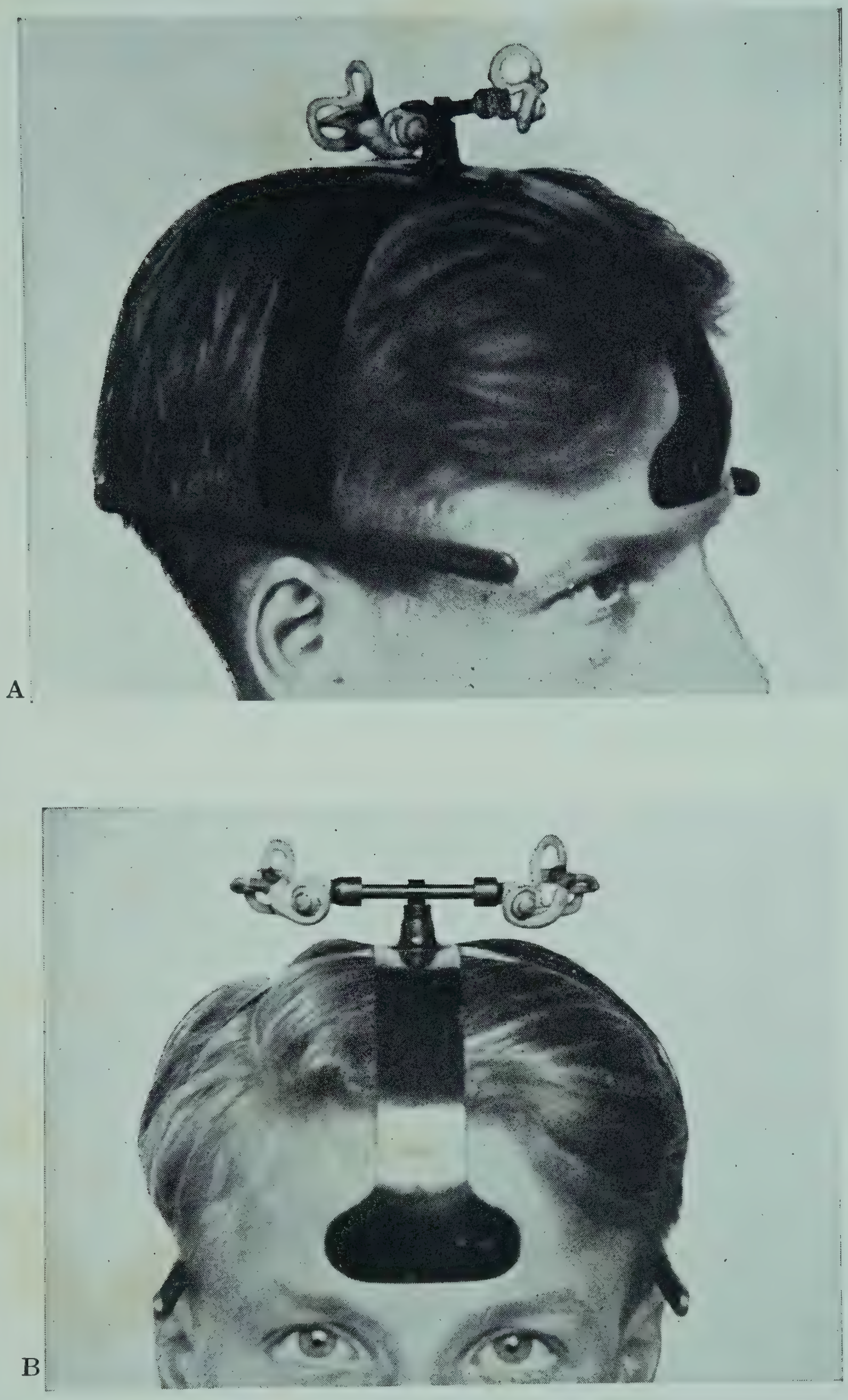
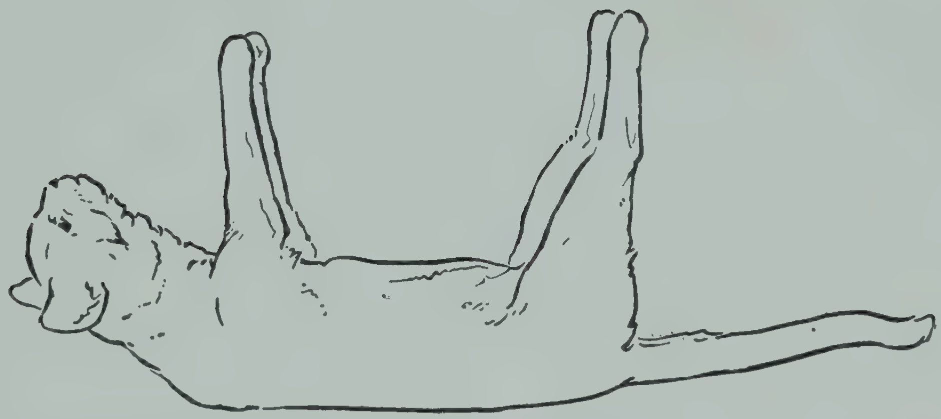
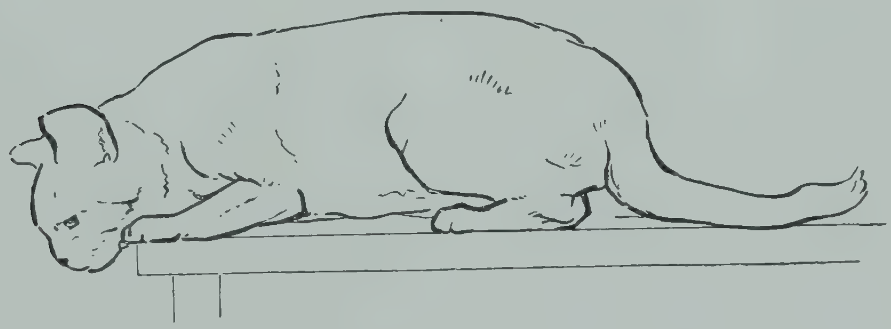


Fig. 46, 18, A and B

Photograph of model prepared under the direction of A. H. Cheatle & V. E. Negus (1927) (*Lancet*, i, 712), to show the relative positions of the cochlea and semicircular canals.



A. Maximum extensor tone (supine position, head 45° up).



B. Minimum extensor tone (prone with head 45° down).

FIG. 46, 19

Tonic labyrinthine reflexes as shown by a decerebrate animal with intact otolith organs.

The semicircular canals can be stimulated in man by rotating the subject in a special chair at a speed of say one revolution in two seconds. If he is rotated to his right both eyes move slowly to the left; this is followed by a quick (saccadic) movement of the eyes to the right which brings the eyes back to the mid position. Another slow and quick movement occurs and so on. This is called vestibular nystagmus. By convention the direction of the nystagmus is taken as the direction of the rapid component. When the rotation to the right is suddenly arrested the sensory impulses from the labyrinths are altered as if the rotation were now to the left, and nystagmus occurs in the opposite direction. The head, body and arms tend to turn in the direction of the rotation just ended. The subject feels that he is still moving (vertigo) and may fall over if he attempts to stand up. This disturbance of equilibrium is probably due to confusion between visual and labyrinthine signals. The duration of the subjective sensation of turning is related to the time taken for the cupola to regain its resting position. The pathway followed by the nerve impulses involved in vestibular nystagmus is from the vestibular nerve to the vestibular area in the brain stem and from there to the nuclei of the eye muscles, higher control being exercised by the temporal lobes.

Excessive stimulation of the semicircular canals, especially if the rotational movement is about a horizontal axis, may produce the drowsiness, pallor, vertigo, salivation, nausea and vomiting characteristic of sea-sickness. Some people suffer in a similar way if they travel on trains or buses or if they go on a swing. Although there may be a large psychological element in the production of their discomfort this is not necessarily the whole explanation because it has recently been shown that removal of the nodulus (posterior lobe) of the cerebellum in dogs gives them complete protection from motion sickness. Experiments made with healthy subjects in swings show that sickness most often occurs if the vertical canals are subjected to acceleration. The supine position (preferably with the eyes open) produces the least motion sickness; in this position stimulation of the vertical canals is at a minimum. Deaf mutes and labyrinthectomized animals show no signs of motion sickness.

The canals may also be stimulated in man by putting either warm or cold water into the external meatus of the ear (caloric tests). The convection currents thus produced in the horizontal canal cause nystagmus, a twisting of the trunk and a feeling of giddiness. Attempts, with the eyes shut, to touch a given point may fail (past-pointing).

These two methods of stimulating the semicircular canals are used in clinical medicine to study the integrity of the labyrinth and its central connexions.

Nystagmus can also be produced by making the subject look at vertical black stripes on a slowly revolving white drum. If the drum revolves so that the black stripes pass to the left the eyes turn slowly to the left to follow a stripe and when this disappears the eyes suddenly move back to the right; the process is repeated with the next black stripe fixated, and so on. The pathway for this response is retina to optic tracts, to anterior corpora quadrigemina, to the nuclei of the eye muscles; the supramarginal and angular gyri exercise a modifying influence on this brain stem pathway. Since this optokinetic nystagmus is not abolished in streptomycin poisoning its pathway must be quite independent of that for vestibular nystagmus.

TONIC LABYRINTHINE REFLEXES

The tonic labyrinthine reflexes can be studied in the decerebrate preparation by cutting the dorsal roots of the upper cervical nerves to abolish the neck reflexes, or by preventing movement of the head and neck with a plaster of Paris bandage. The tonic labyrinthine reflexes, unlike the tonic neck reflexes (p. 855), produce similar alterations of tone in all the four limbs. When the head is in the position of Fig. 46, 19, A, the otoliths pull on the utricular maculæ and extensor tone is at a maximum, but if the animal is in the position shown in Fig. 46, 19, B, so that the head points down 45°, the extensor tone is at a minimum. These reflexes disappear when the otolith organs are destroyed as may be done, without damage to the semicircular canals, by centrifuging the animal. The tonic labyrinthine reflexes frequently modify the tonic neck reflexes already described. For example, depressing the nose reduces the tone in the fore-limbs through both reflex mechanisms, whereas when the nose is elevated the extensor tone in the hind-limbs is decreased by the neck reflexes and increased by the labyrinthine reflexes.

RIGHTING REFLEXES

The animal with the thalamus and brain stem intact but with the cerebral cortex and cerebellum removed, 'the thalamic animal,' possesses still more adequate postural reflexes. The thalamic cat or dog shows normal posture and it can carry out walking movements. The muscle tone is normal when sitting or walking but if the animal is supported with the feet off the ground hyperextension of all four limbs occurs.

The thalamic animal shows positive supporting reactions which convert the limbs into pillars by simultaneous contraction of the flexors and extensors so that the body weight is taken on four columns. The actual contact of the feet with the ground produces an extensor response and at the same time the flexor muscles in the distal joints of the feet are stretched. This causes reflexly a contraction of all the muscles, extensors, flexors, abductors and adductors, which allows the limbs to carry the weight of the body. At the same time the vertebral column is made rigid by a contraction of the long muscles which run dorsally and ventrally along it. This supporting reaction is a static reflex and necessarily disappears when movements are carried out; the limb loosens (negative supporting reaction) when the extensor muscles of the distal joints are stretched. These supporting reactions, although shown quite well by the thalamic animal, are best seen in the intact animal.

In contradistinction to the animal with decerebrate rigidity, the thalamic animal possesses righting reflexes which ensure that the animal brings itself back to the upright position if it is made to deviate from it. The righting reflexes arise in the labyrinths and in the muscles of the neck and trunk. To study the components of the righting mechanism it is necessary to employ one or more of the following procedures—labyrinthectomy, denervation of the neck, blindfolding. When the results of such experiments are put together it is possible to explain how an animal rights itself when put in an abnormal position. If a cat is held upside down above the ground and released, it turns over in mid-air and lands on all four feet. In this manœuvre probably the eye righting reflexes are the most important; they are dependent on the integrity of the visual area of the cortex in the occipital lobe. The animal, when it sees its surroundings, quickly brings its head to the normal position relative to the ground. The labyrinthine reflexes are also of some importance in correcting the position of the head since the otoliths inform the animal of the position of its head in space before it is released but a cat with congenital absence of the labyrinths (provided it was not blindfolded) could turn over without delay quite nimbly in mid air. A normal cat blindfolded and then released turns over after a delay of about 100 msec but makes a poor landing as it has no information about the position of the floor. A recent reinvestigation of this old problem by McDonald who used high-speed (1500 frames per sec) cinephotography has shown that the rotation after release occurs in two principal stages (1) the righting of the head and fore-part of the body together and (2) the rotation of the lumbar region, pelvic girdle and hind-limbs all in one, the whole turn taking less than 150 msec.

A further source of righting is displayed by a blindfolded labyrinthectomized animal laid on its side on a table. When the animal is released the head immediately rotates into the normal position relative to the ground and the position of the remainder of the body is corrected as before. If the animal is laid on its side again under a weighted board, so that the pressure on the body is the same on both sides, this righting reflex does not occur.

The righting reflexes which originate in the labyrinths and in the neck muscles are integrated at the level of the red nucleus, probably in the reticular substance around the red nucleus. Removal of the cerebellum does not destroy them. The optical righting reflexes, on the other hand, are dependent on the integrity of the cerebral cortex.

When the chin of an intact blindfolded cat is laid on a table, or when the fore-legs are brought into contact with the edge of the table, the feet are immediately placed on the table in a position ready to support the weight of the body. These are called *placing reactions*. When an intact animal is disturbed by being pushed sideways the normal position is regained by a series of hops (*hopping reaction*). The hopping and placing reactions are dependent on the integrity of the motor cortex.

The precise details of the effect on posture of section at any given level in the brain vary according to the complexity of the nervous system. In a primate with the brain stem cut between the superior and inferior quadrigeminal bodies, the fore-limbs are rigidly extended and thrust backward. In a high decerebration in a primate, better described as a decortication, the fore-limbs are flexed, while the legs are extended but the animal cannot walk. Decorticate rigidity is occasionally seen in man as a terminal phenomenon when the cortex is extensively damaged (Fig. 46, 20); the patient lies on his back with arms flexed and legs extended and adducted. Rotation of the head causes alteration in the tone of the arms as described in dealing with tonic neck reflexes (p. 855). Lengthening and shortening reactions (p. 843) can be obtained, so that the rigidity is 'plastic' as in lower animals.

CONSCIOUSNESS

It is often assumed that, since the cerebral cortex carries out the intellectual functions, consciousness or awareness of our surroundings depends on that part of the brain. Purely cortical lesions, however, do not produce unconsciousness. Unconsciousness is more likely to occur in lesions of the upper brain stem and diencephalon.

REFERENCES

ADRIAN, E. D. (1943). Discharges from the vestibular receptors in the cat. *J. Physiol.* **101,** 389–407.

BRAIN, W. R. (1950). Cerebral basis of consciousness. *Brain,* **73,** 465–479.

CARMICHAEL, E. A., DIX, M. R. & HALLPIKE, C. S. (1954). Lesions of the cerebral hemispheres and their effects on optokinetic and caloric nystagmus. *Brain,* **77,** 345–372.

DELAFRESNAYE, J. F. (Ed.) (1954). *Brain Mechanisms and Consciousness. A Symposium.* Oxford: Blackwell.

VAN EGMOND, A. A. J., GROEN, J. J. & JONKEES, L. B. W. (1949). The mechanics of the semicircular canal. *J. Physiol.* **110,** 1–17.

GERNANDT, B. (1949). Responses of mammalian vestibular neurons to horizontal rotation and caloric stimulation. *J. Neurophysiol.* **12,** 173–184.

GRANIT, R., HOLMGREN, B. & MERTON, P.A. (1955). The two routes for excitation of muscle and their subservience to the cerebellum. *J. Physiol.* **130,** 213–224.

LOEWENSTEIN, O. (1936). Equilibrium function of the vertebrate labyrinth. *Biol. Rev.* **11,** 113–145.

LOEWENSTEIN, O. & SAND, A. (1940). Activity of the semicircular canals. *J. Physiol.* **99,** 89–101.

MCDONALD, D. A. (1955). How does a falling cat turn over? *St Bart's. Hosp. J.* **56,** 254–258.

MAGOUN, H. W. (1950). Caudal and cephalic influences of the brain stem reticular formation. *Physiol. Rev.* **30,** 459–474.

TYLER, D. B. & BARD, P. (1949). Motion sickness. *Physiol. Rev.* **29,** 311–369.

YOUNG, W. (1955). Étude experimentale du drainage labyrinthique de la périlymphe. VI Congrès Fédératif international d'Anatomie.

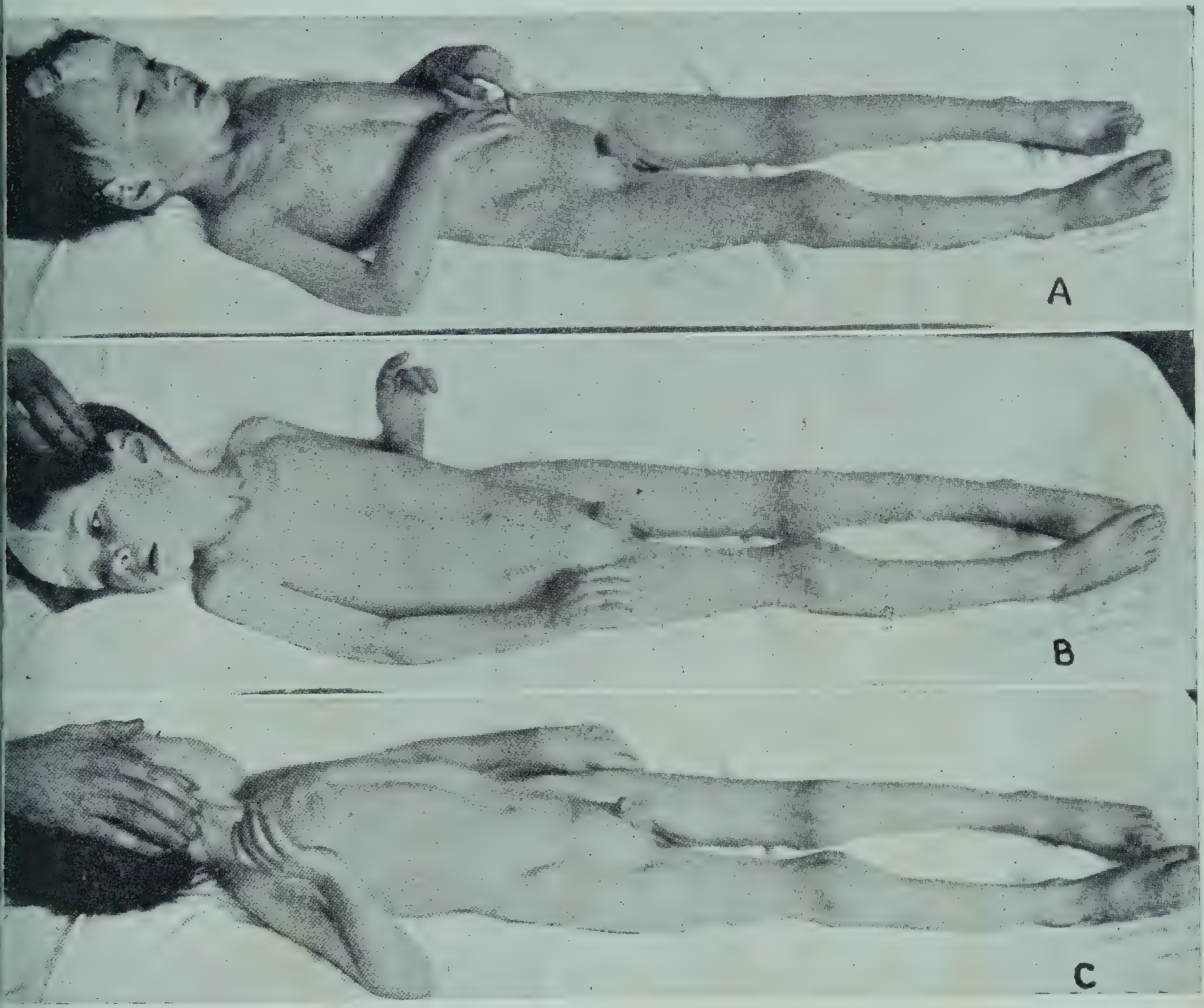


Fig. 46, 20

Neck reflexes shown in decerebrate man. A, the characteristic position, flexion of the arms with xtension of the legs. B and C, the position of the neck affects the tone of the muscles of the arms (L. E. Davis (1925), *Arch. Neurol. Psychiat. Chicago*, **13,** 572).

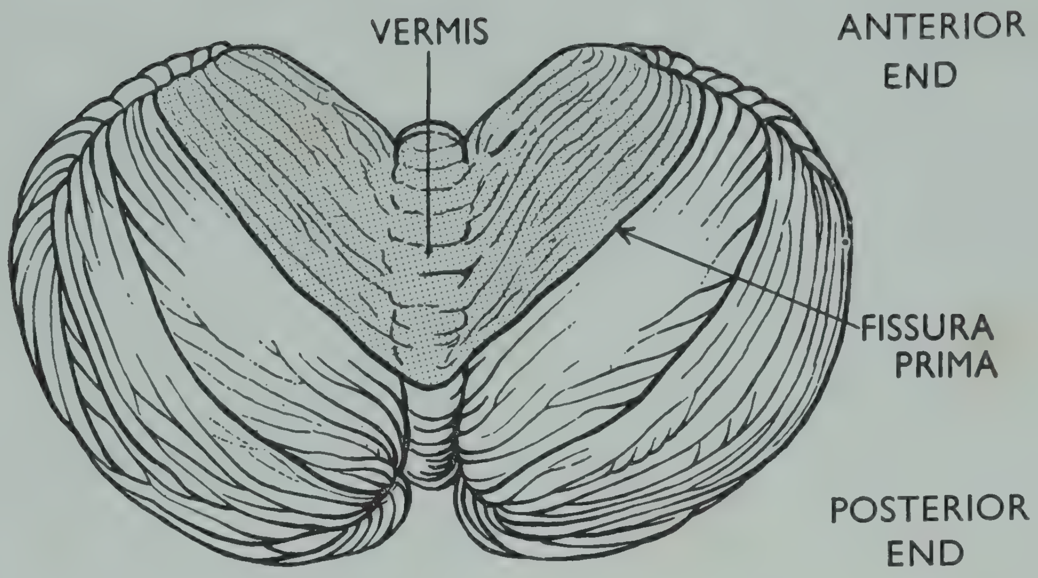

VERMIS
ANTERIOR END
FISSURA PRIMA
POSTERIOR END

Fig. 47, 1

The superior aspect of the cerebellum. The anterior lobe is stippled. The remainder belongs to the middle lobe. (*After Gray's Anatomy* (1954), 31st ed., p. 961. London: Longmans.)

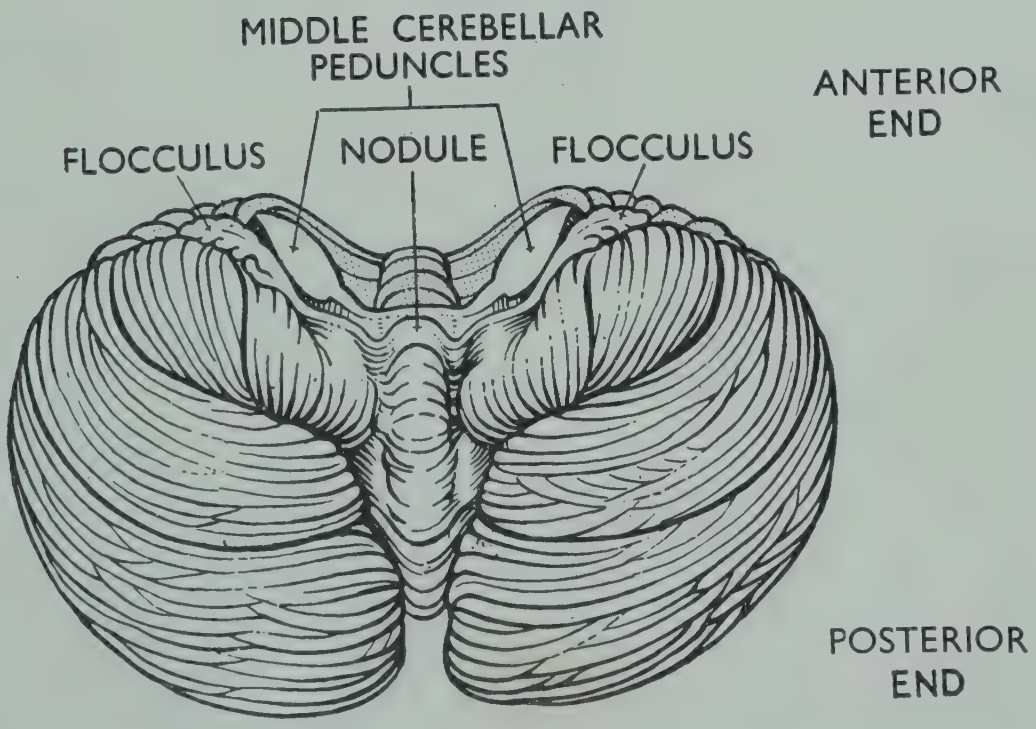

MIDDLE CEREBELLAR PEDUNCLES
ANTERIOR END
FLOCCULUS
NODULE
FLOCCULUS
POSTERIOR END

Fig. 47, 2

Inferior aspect of the human cerebellum. The flocculus and nodule (boldly outlined) form the posterior lobe. The anterior lobe is stippled. The remainder belongs to the middle lobe. (*After Gray's Anatomy* (1954), 31st ed. p. 964. London: Longmans.)

CHAPTER 47

THE CEREBELLUM

THE cerebellum, like the cerebral hemispheres, consists of a superficial layer of grey matter, the cortex, overlying a mass of white matter. The cerebellum has a large number of fine parallel furrows on its surface (Fig. 47, 1 and 2) and a section through it shows a tree-like arrangement of grey and white matter, the *arbor vitæ.*

The cerebellum is developed from the cephalic portion of the hind-brain and in man can be divided on embryological and phylogenetic as well as on functional grounds into anterior, middle and posterior lobes. These lobes are demarcated by fissures which are evident at the third or fourth month of embryonic life (Fig. 47, 3). The relative sizes of the three lobes in the adult are shown in Figs. 47, 1 and 2, and their main afferent connexions are given in Table 47, 4. The anterior and posterior lobes are much older in a phylogenetic sense than the middle lobe and are, therefore, called *palæocerebellar.* The great development of the middle lobe (*neocerebellum*) of the cerebellum in man has caused an alteration in the anatomical relations of the lobes as seen in the embryo, so that the posterior lobe eventually lies below and in front of the middle lobe.

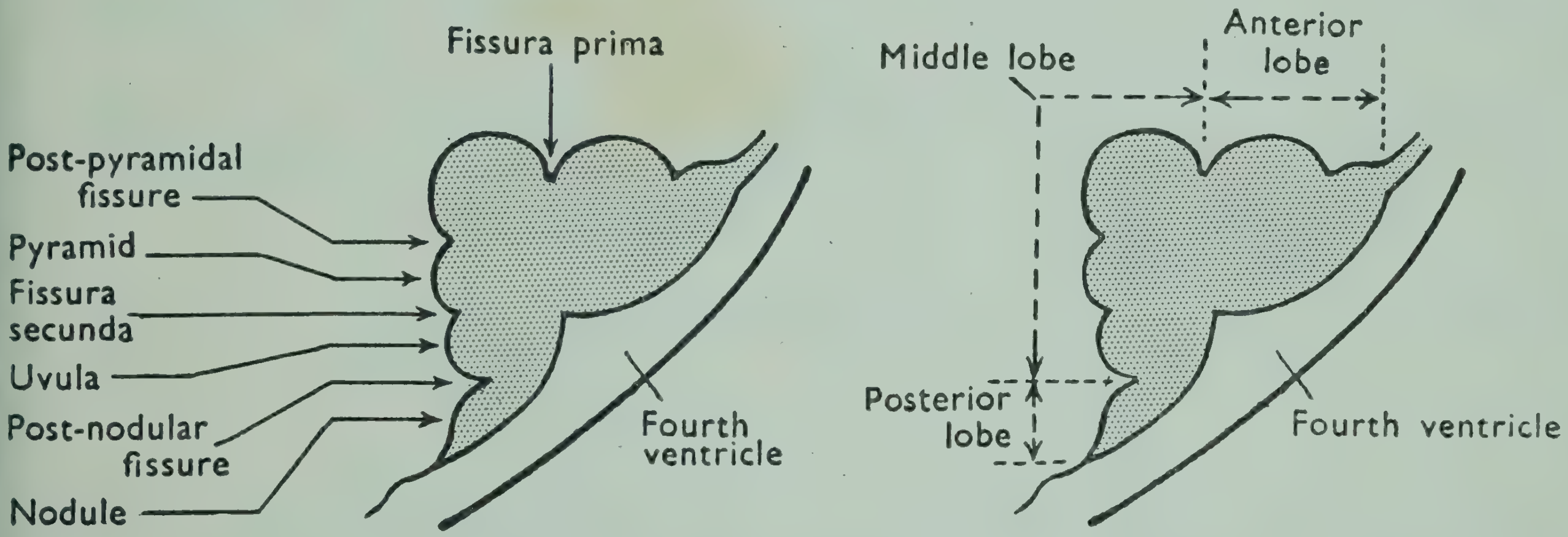


FIG. 47, 3

A diagrammatic representation of a saggital section through the cerebellum in a human four-months fœtus. (*After* Gray.)

The three cerebellar peduncles carry projection fibres connecting the cerebellum with other parts of the central nervous system. The majority of the fibres in the *superior cerebellar peduncles* pass from the cerebellar nuclei through the mid-brain where they decussate (Fig. 47, 6) and then divide into descending branches to the upper brain stem and ascending branches which go to the red nuclei and to the thalami. The superior peduncles also convey the anterior spinocerebellar fibres which end mainly in the cortex of the anterior lobe. The *middle peduncles,* the largest afferent connexions of the cerebellum, carry impulses from the cerebral cortex. These travel to the cells of the nuclei pontis from which new axons, crossing the mid-line in the pons, pass to the cortex of the middle lobe of the cerebellum of the opposite side. The *inferior peduncles* carry fibres from various brain stem nuclei and from the spinal cord as well as certain efferent fibres. The chief fibres are (1) the posterior spinocerebellar tract which ascends without crossing and ends in the pyramid and uvula (the palæocerebellar parts of the middle lobe); some fibres of this tract also go to the anterior lobe; (2) arcuate fibres from the lateral cuneate nuclei of both sides passing to an unknown destination; (3) fibres from the vestibular nuclei and from the vestibular nerve passing to the cortex of the posterior lobe; (4) fibres passing from the cortex of the posterior lobe and from the nucleus fastigii to the vestibular nucleus in the medulla oblongata; (5) olivocerebellar fibres passing via the inferior cerebellar peduncle to the vermis and cerebellar hemisphere of the opposite side.

TABLE 47, 4

Lobes of the Cerebellum in Man

Lobe	Separating fissure	Phylogeny	Relative size	Main afferent connexions
Anterior		Palæocerebellum	Small	Incoming spinocerebellar
	Fissura prima			
Middle		Neocerebellum	Large	Corticoponticerebellar
	Postnodular and floccular			
Posterior		Palæocerebellum	Very small: nodule and flocculi only	Vestibular

The caudal parts of the middle lobe (uvula and pyramid), which receive the posterior spinocerebellar fibres, belong to the palæocerebellum. It should be noted that there are several other schemes of classification of the parts of the cerebellum. In one such classification the posterior lobe comprises not only the nodulus and the flocculus but also the uvula and pyramid.

The grey matter on the surface of the cerebellum has a very characteristic histological appearance owing to the presence in it of the large Purkinje cells, but there are no local differences, such as exist in the cerebrum, sufficient to distinguish one part of the cerebellar cortex from another. The appearance of the cells and their connexions are given in Fig. 47, 5. The Purkinje cells are large flask-shaped cells arranged in a single layer; each has an arborescent arrangement of dendrites and a single axon which passes down to one of the deep nuclei of the cerebellum. The *climbing fibres* make synaptic connexions with the dendrites of the Purkinje cells; these afferent fibres may be the terminations of fibres from the vestibular and pontine nuclei. The Purkinje dendrites are also associated with dendrites from *basket cells* and *granular cells.* The granular cells receive the terminations of *moss* fibres which may be the main afferent fibres of the cerebellum. Practically all the afferent fibres of the cerebellum terminate in the cortex.

Independent centres of grey matter, four on each side, lying in the white matter of the cerebellum receive the axons of the cells of Purkinje and are the source of nearly all the efferent fibres of the cerebellum. The nucleus dentatus, the largest and phylogenetically the youngest of the nuclei, is connected with the neocerebellum and through the superior cerebellar peduncles with the thalamus and cerebral cortex (areas 4 and 6) of the opposite side. The dentate nucleus also sends fibres across the mid-line to the red nucleus from which new fibres cross the mid-line and descend as the rubrospinal tract. The other, smaller, nuclei are the nucleus fastigii, the nucleus globosus and the nucleus emboliformis; the nucleus fastigii sends fibres to the vestibular nucleus and to the reticular formation of the medulla. The globose and emboliform nuclei are most likely to be connected by the superior peduncles to the large cells of the red nucleus and the reticular formation of the mid-brain.

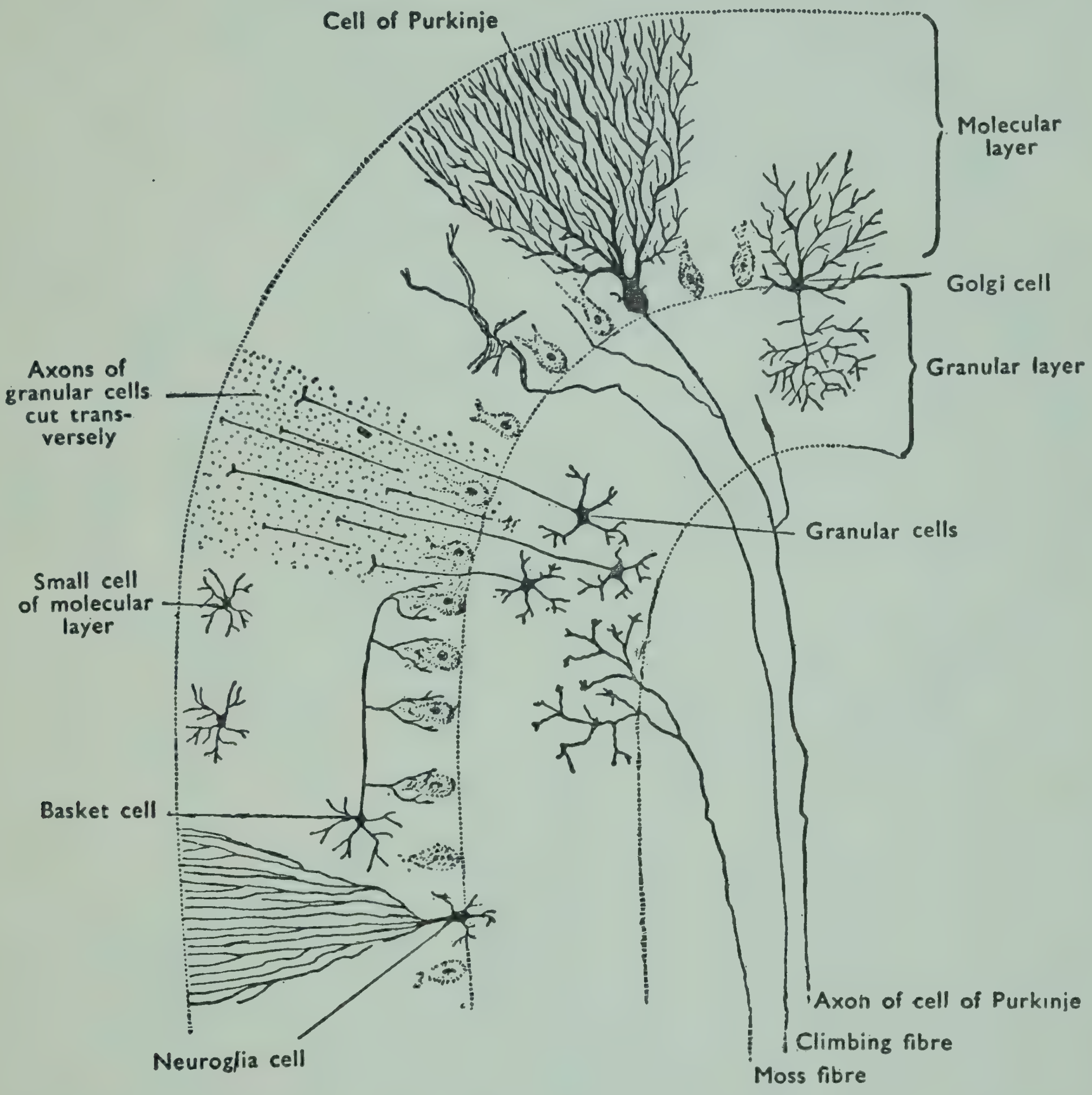

Cell of Purkinje
Molecular layer
Golgi cell
Granular layer
Axons of granular cells cut transversely
Granular cells
Small cell of molecular layer
Basket cell
Neuroglia cell
Axon of cell of Purkinje
Climbing fibre
Moss fibre

FIG. 47, 5

A transverse section through a cerebellar folium (*Gray's Anatomy* (1954), 31st ed., p. 969. London: Longmans). (*After* Cajal & Kölliker.)

By the use of electrical methods Adrian has recently amplified our knowledge of the connexions of the cerebellum. Under barbiturate (dial) anæsthesia a fine wire was inserted about 1·5 mm deep into the cerebellum of a monkey and afferent impulses were picked up when the limbs on the same side were pressed or moved (Fig. 47, 7). A movement of the vibrissæ on the face produced a discharge in the anterior part of the middle lobe.

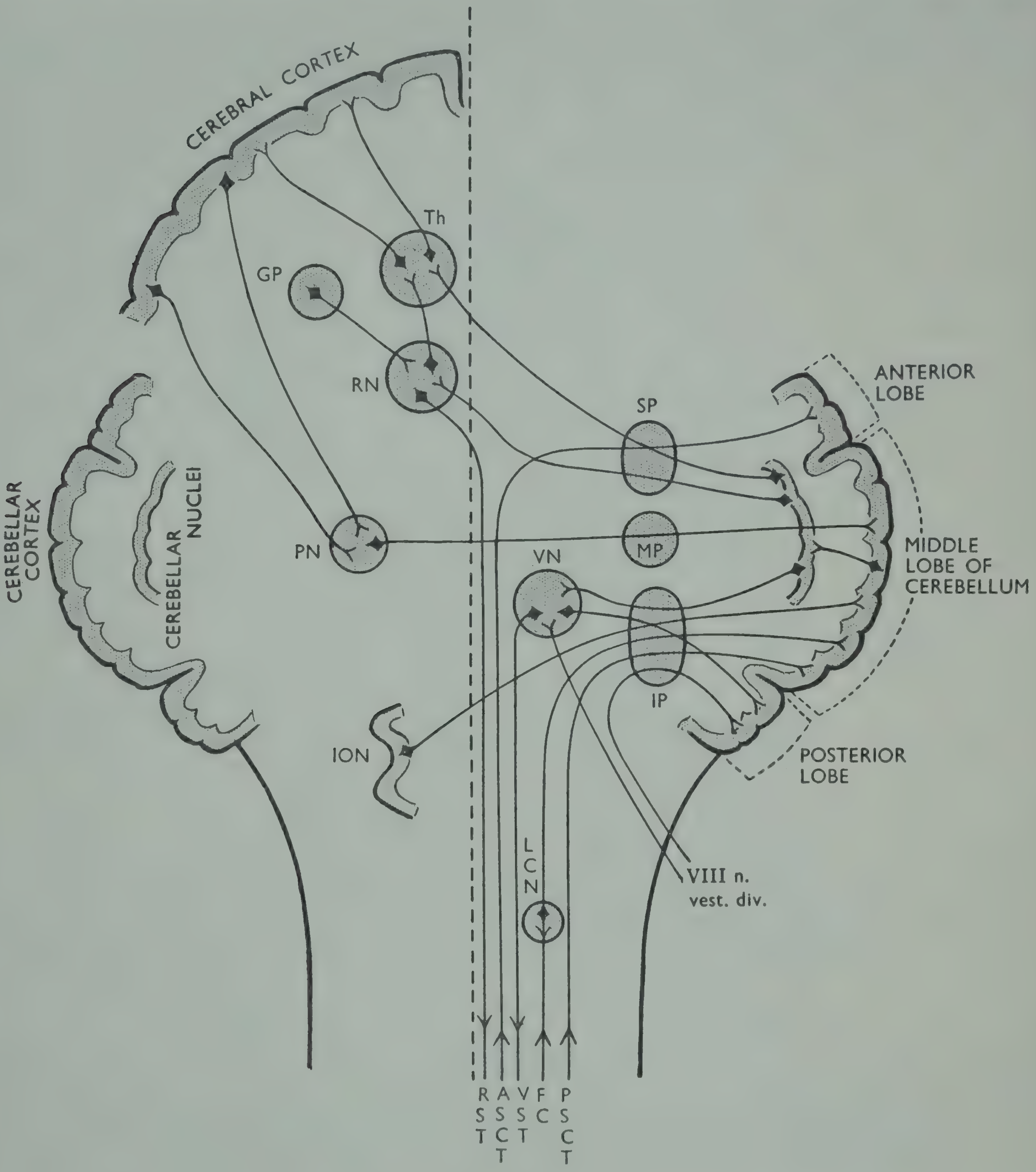

CEREBRAL CORTEX
Th
GP
RN
PN
CEREBELLAR CORTEX
CEREBELLAR NUCLEI
SP
MP
VN
IP
ION
ANTERIOR LOBE
MIDDLE LOBE OF CEREBELLUM
POSTERIOR LOBE
LCN
VIII n. vest. div.
RST
ASCT
VST
FC
PSCT

FIG. 47, 6

A simplified scheme of the connexions of the cerebellum. Note that each side of the cerebellum is connected mainly with the same side of the spinal cord and with the opposite cerebral cortex.

ASCT = anterior spinocerebellar tract.
LCN = lateral cuneate nucleus.
FC = fasciculus cuneatus.
GP = globus pallidus.
IP = inferior peduncle.
MP = middle peduncle.
ION = inferior olivary nucleus.
PN = pontine nuclei.
PSCT = posterior spinocerebellar tract.
RN = red nucleus.
RST = rubrospinal tract.
SP = superior peduncle.
Th = thalamus.
VN = vestibular nuclei.
VST = vestibulospinal tract.
VIII n. = eighth cranial nerve.

Dorsiflexion of the wrist of the fore-limb or pressure on the hand produced discharges just craniad to the fissura prima, while pressure on the foot of the hind-limbs gave discharges well forward. Some afferent impulses to the cerebellum arose from the surface of the body but the majority came from the muscles, tendons and fasciæ of hand and foot. These experiments show that spinocerebellar afferents end chiefly in the anterior lobe. The rhythmical changes of electrical potential characteristic of the cerebral cortex can also be picked up in the middle cerebellar peduncles and over a large area of the anterior and middle lobes of the cerebellum. When the cerebral cortex is stimulated, either electrically or by application of a solution of strychnine, discharges are obtained in the cerebellum in the contralateral areas shown in Fig. 47, 8. Thus the cerebropontine impulses reach cerebellar areas in which the foot is represented craniad and the hand caudad. This is the same order of distribution as in the case of the afferent spinocerebellar fibres but the cerebropontine afferents reach more caudad into the middle lobe. Snider and Stowell also studied the distribution of the spinal projection to the anterior lobe and in addition found visual and auditory receiving areas largely in the middle lobe, the visual and auditory areas being almost coextensive. The activity of the cerebellar cortex itself is demonstrated by laying electrodes on its surface; it shows small waves of high frequency (150 to 250 per sec) which are increased in size and frequency when an afferent discharge reaches the cerebellum.

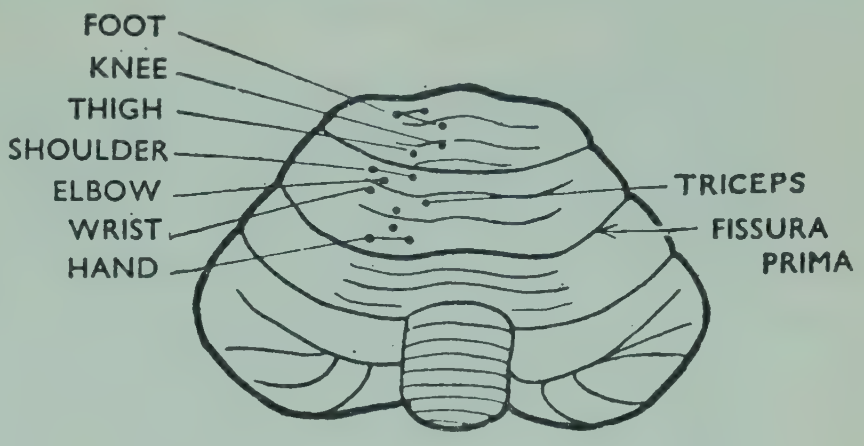


Fig. 47, 7

Localization of discharges in the dorsal surface of the cerebellum from different parts of the fore- and hind-limb in a monkey (E. D. Adrian (1943), *Brain*, **66**, 298).

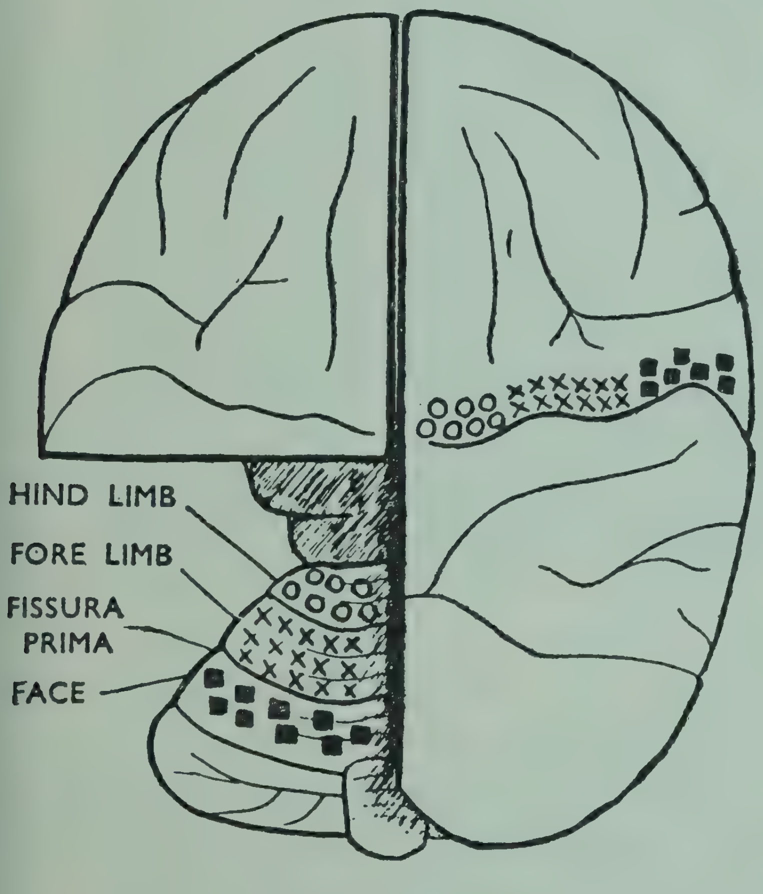


Fig. 47, 8

Subdivisions of pontocerebellar receiving area in the monkey, showing their connexion with different parts of the motor cortex. Stimulation of the area of the motor cortex marked ○ (leg area) produced potentials in the cerebellum in the area marked ○; a similar relationship held for the areas marked **x** and ■ (E. D. Adrian (1943), *Brain*, **66**, 301).

Extirpation (ablation) of parts of the central nervous system, although a useful method for investigating function, has given results which must be interpreted with care in the case of the cerebellum because it is difficult to remove this part of the brain without damaging the adjacent brain stem. Since the cerebellum possesses a modifying rather than an initiating function its removal does not destroy any reflexes although it usually alters their character (Fig. 47, 10). It is important to note that in no animal is there any loss or disturbance of sensation after removal of the cerebellum; the gait is abnormal after decerebellation not because there is any disorder of the sensory pathways but because the muscles are not under normal control.

More than fifty years ago Luciani studied the effects of removal of the cerebellum. He found that in dogs the operation was followed by an increase in extensor tone and by clonic spasms. These signs passed off and in about ten days the decerebellate dog showed *asthenia* (weakness), *hypotonia* (reduction of the tone of the muscles) and irregular contractions of muscles instead of the normal smooth, well-regulated muscle movements. This phase was followed by a certain amount of compensation in which standing and walking were much improved but recovery was never complete (Fig. 47, 9). The improvement is due largely to the use of the eyes, for if the eyes of an animal which has recovered from the initial effects of decerebellation are covered the gait at once deteriorates (Fig. 47, 9, d). Luciani also found, as have all subsequent investigators, that removal of one lateral half of the cerebellum is followed by weakness and defects of muscular control on the same side of the body.

Later investigators, with improved technique, have shown that, provided the vestibular nuclei are unharmed, asthenia and hypotonia are not seen in decerebellate dogs or cats. When the period of extensor spasm has passed off the animal is unable to carry out complicated reflexes such as those involved in locomotion in the normal way (see Fig. 47, 9), but simpler, essentially spinal, reflexes such as scratching may not be greatly altered.

In healthy animals and healthy human beings a small amount of tremor is always present although the movements are so slight as to escape observation. This tremor consists of small adjustments of flexion and extension, or of abduction and adduction, which are needed to maintain body posture or the position of a limb, especially if it is unsupported. In a decerebellate animal this tremor is greatly exaggerated and is especially obvious when movements are being carried out; the synergic muscles do not act harmoniously. The tremor together with the hypotonia and asthenia gives the jerky staggering gait described as *cerebellar ataxia*.

The postural reflexes of the decerebellate dog are usually exaggerated but the labyrinthine and other righting reactions are such effective compensating mechanisms that the animal maintains its equilibrium in spite of the ataxia. Rademaker described such an animal as standing with its feet wide apart, with the legs hyperextended because of an exaggeration of the positive supporting reaction and with the vertebral column convex dorsally. The overactivity of the righting reactions seems to be a major factor in producing ataxia, since any error in stance or balance is overcorrected and this exaggerated movement leads to another overcorrection so that progression tends to be somewhat zigzag. Jerking movements of the eyes, described as nystagmus, are not seen in decerebellate dogs provided that the vestibular nuclei are undamaged.

In higher animals the results of decerebellation are somewhat different. If the whole cerebellum is removed in the monkey the animal shows at first extensor spasms which are not as marked as in the dog; the ataxia which follows is so gross that it is some weeks before the animal has recovered sufficiently to be able to walk. The animal has a marked *intention tremor* when it attempts to carry out any voluntary movements. Hypotonia may be present but the postural reflexes are not overactive as in the dog.

The cerebellum possesses localization of function to a limited extent and partial removal shows that different parts vary in their regulatory functions.

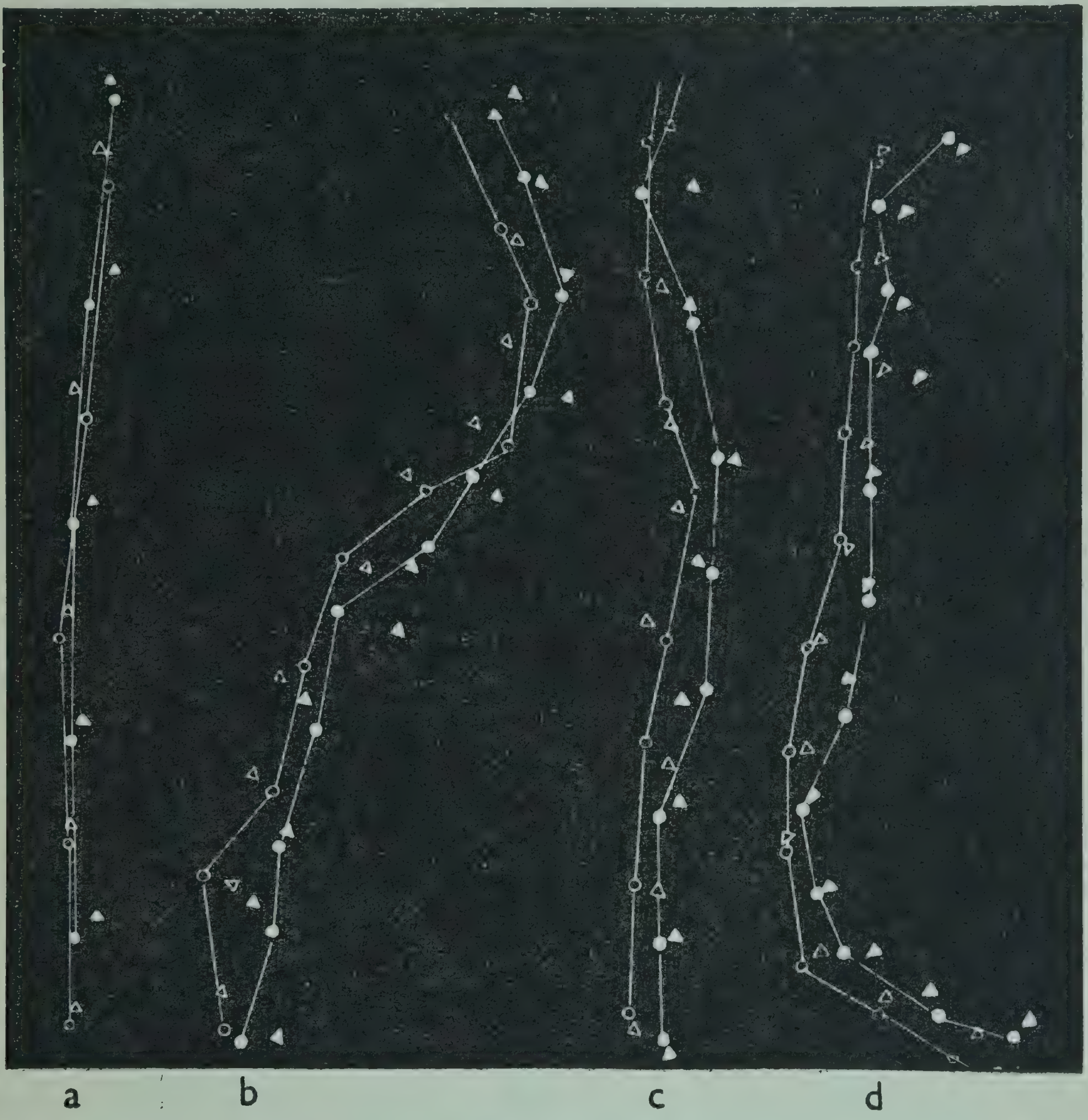


FIG. 47, 9

Tracing of the footprints during progression in the dog. The prints of the fore-legs are represented by small circles, those of the hind-legs by triangles. The prints of the left leg are distinguished from those of the right leg by a black space in the centre of the circles and triangles. The traces of the right and left feet are united, respectively, by lines. Each 25 mm corresponds to 1 m. a, shows the gait of a normal bitch weighing 6 kg. b, c, d, show the gait of a bitch weighing 6 kg. in which the cerebellum had been almost completely removed by three operations. b, tracing obtained two and a half months after final operation, c, eleven months after, d, eleven months after, with the eyes blindfolded (L. Luciani (1915), *Human Physiology*, vol. iii, pp. 438 and 449. London: Macmillan).

Each half (right or left) of the cerebellum modifies the tone and movements of the muscles on the same side of the body (see Fig. 47, 6). If only the flocculonodular (posterior) lobe is removed in monkeys there is an oscillation of the head, a tendency to fall, and the animal walks with a staggering gait with the limbs abducted. This is due to overactivity of the vestibular apparatus since it does not occur if the labyrinths are previously destroyed. There is, however, no tremor and no change in reflexes. If the inferior peduncles are cut the results are similar. Lesions of the posterior lobe, seen occasionally in children, lead to great difficulty in maintaining balance; the feet are kept wide apart and there is a tendency to fall backwards but there is no tremor. If the anterior lobe is removed in monkeys the stretch reflexes and positive supporting reactions are augmented. Since the middle lobe with its cortical and thalamic connexions is especially developed in the higher mammals it is not surprising to find that removal of the cerebellar cortex in primates produces a more profound disturbance of function than it does in dogs and cats. Removal of the neocerebellar cortex in monkeys causes hypotonia and interferes with skilled movements, but tremor is not seen unless the underlying dentate nucleus is also removed. Removal of neocerebellar cortex together with the dentate nucleus disturbs the gait; the animal may leap and be unable to arrest its progress when confronted with an obstacle. Removal of one side of the neocerebellum in monkeys causes hypotonia of the muscles of the same side. Monkeys show considerable recovery from these lesions but there is much less recovery in chimpanzees and man.

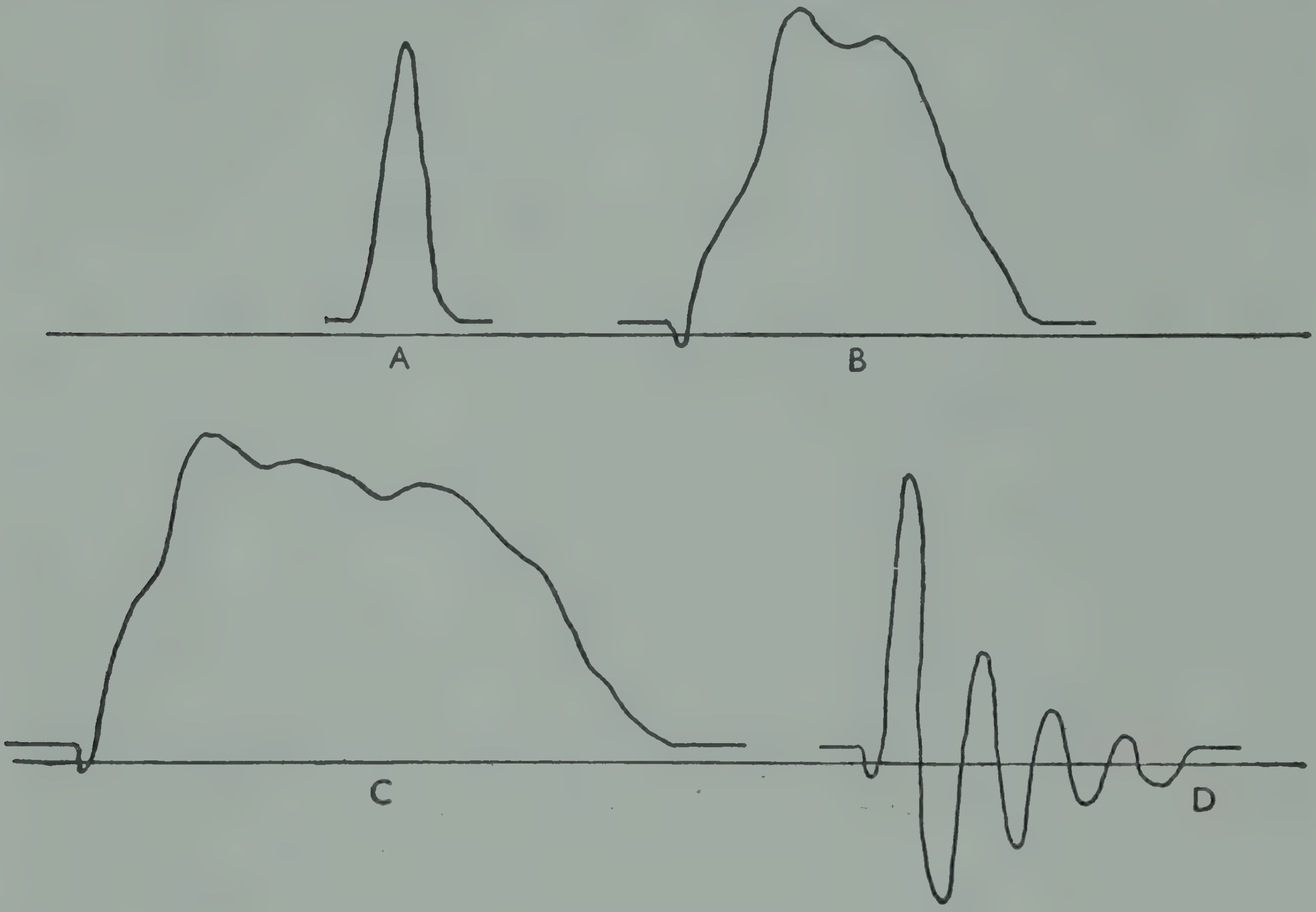


FIG. 47, 10

A—tracing of a simple muscle twitch, B—of a normal knee jerk, C—of a knee jerk in a spastic limb, D—of a pendular knee jerk from a patient with an acute cerebellar lesion (Gordon Holmes (1952), *Introduction to Clinical Neurology*, 2nd ed., p. 100. Edinburgh: Livingstone).

The tremor seen after removal of the cerebellum is called 'intention' tremor since it is not usually present at rest but becomes obvious when voluntary movements are attempted. In a decerebellate animal removal of the motor cortex abolishes the tremor; it also abolishes all voluntary movement but not, of course, reflex movements of the limbs. Thus the tremor observed after removal of the cerebellum alone depends on the loss of the modifying influence of the cerebellum on the motor cortex. In the monkey it has been found that after the removal of the premotor cortex (area 6) on the side opposite to that from which the cerebellum has been removed the tremor is increased. Presumably, therefore, the premotor area plays an important part in the phase of compensation after a cerebellar lesion. The extensor spasms which follow decerebellation are presumably due to a release from cerebellar control. The increase of decerebrate rigidity which follows removal of the cerebellum or of the anterior lobe alone has been discussed on p. 855; as a corollary to this, electrical stimulation of the anterior lobe decreases the extensor tone of decerebrate rigidity. These observations indicate that in the intact animal the cerebellum must exercise an important co-ordinating influence on higher and lower parts of the nervous system, probably through connexions with the reticular formation of the brain stem (p. 850).

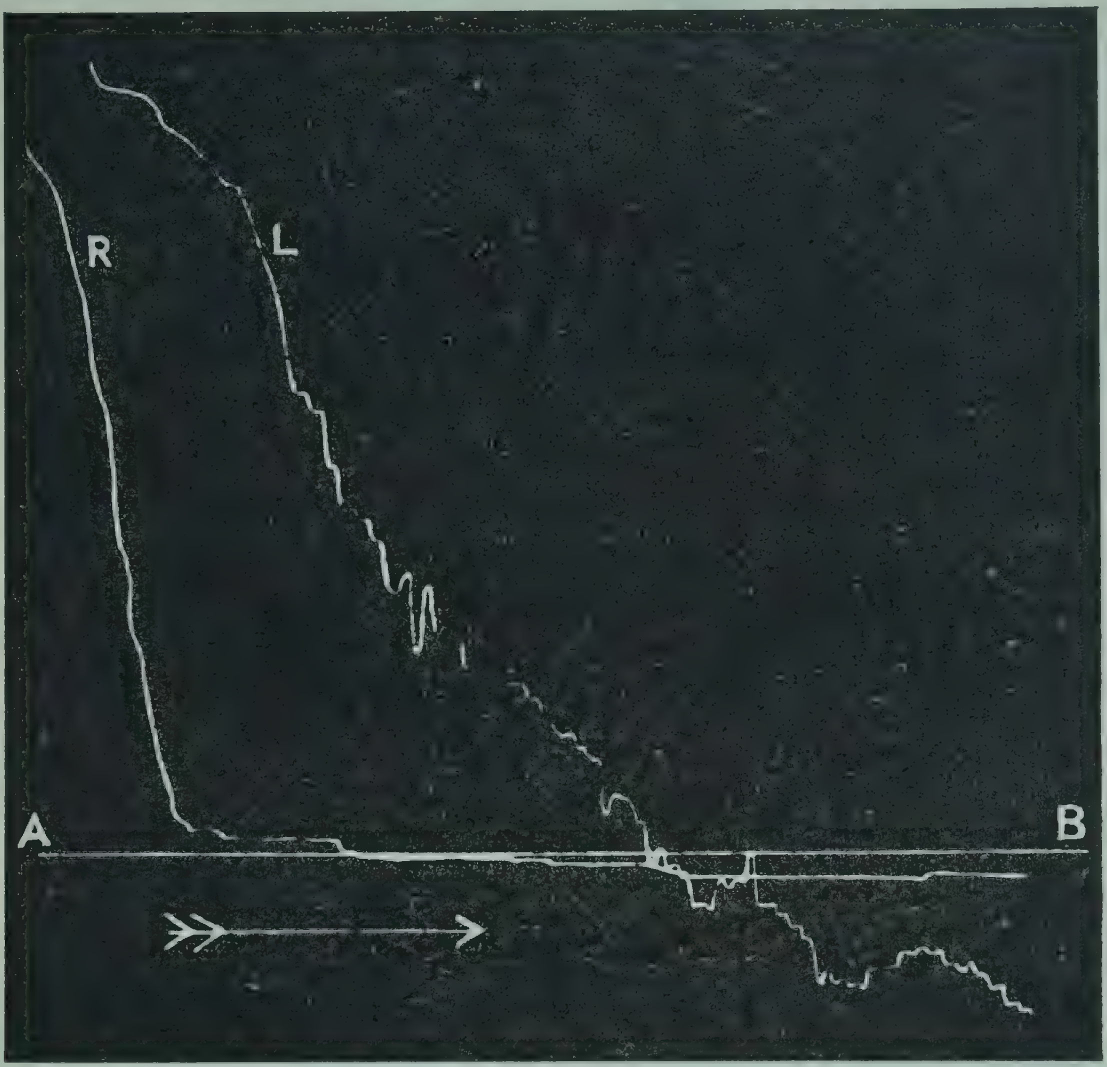


Fig. 47, 11

The patient, who had a wound of the left side of his cerebellum, was asked to extend springs of equal strength with his two arms on a signal and to maintain the stretch at a level indicated by the line A, B. Note the slowness of the left arm in exerting power, the irregular tremulous character of the movement and the failure to maintain the final posture. The record was made on a slowly revolving kymograph. Read from left to right (Gordon Holmes (1939), *Brain*, **62**, 13).

The neocerebellum, because of its size and superficial position, is the part most often damaged in man and, therefore, more is known of the effect of lesions in it than of lesions in the palæocerebellum which is more deeply placed. In gunshot wounds of the cerebellum, in which the vestibular nuclei do not appear to be involved, hypotonia with asthenia and disturbances of voluntary movement are the main effects (Fig. 47, 11). In unilateral lesions these effects are confined to the same side of the body as the lesion. Unilateral cerebellar injury in man never causes disturbances in only one ipsilateral limb and thus both the arm and the leg on the

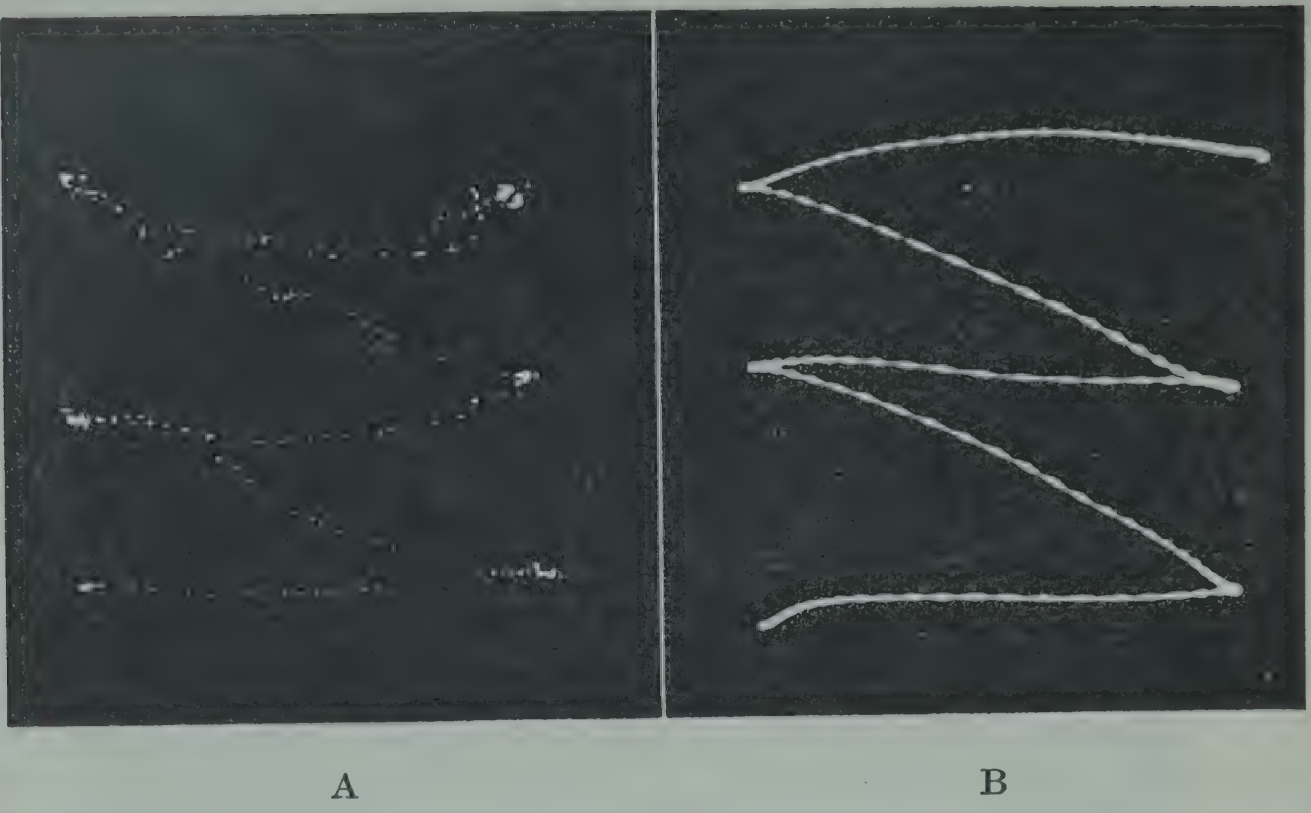


Fig. 47, 12

Records obtained by photographing points of light attached to the tips of the forefingers as a patient with a lesion of the right side of the cerebellum attempted to move each finger slowly and accurately between series of luminous red points (not visible in the photographs) in a dark room. Each flash of light corresponds to 0·04 second. The range of each movement was about 75 cm. B is the record obtained from the left hand, A, that from the right. The irregularity in the rate and in the directions of movements of the affected hand, and the failure to arrest this finger accurately at the points, are well shown in A (Gordon Holmes (1939), *Brain*, **62**, 12).

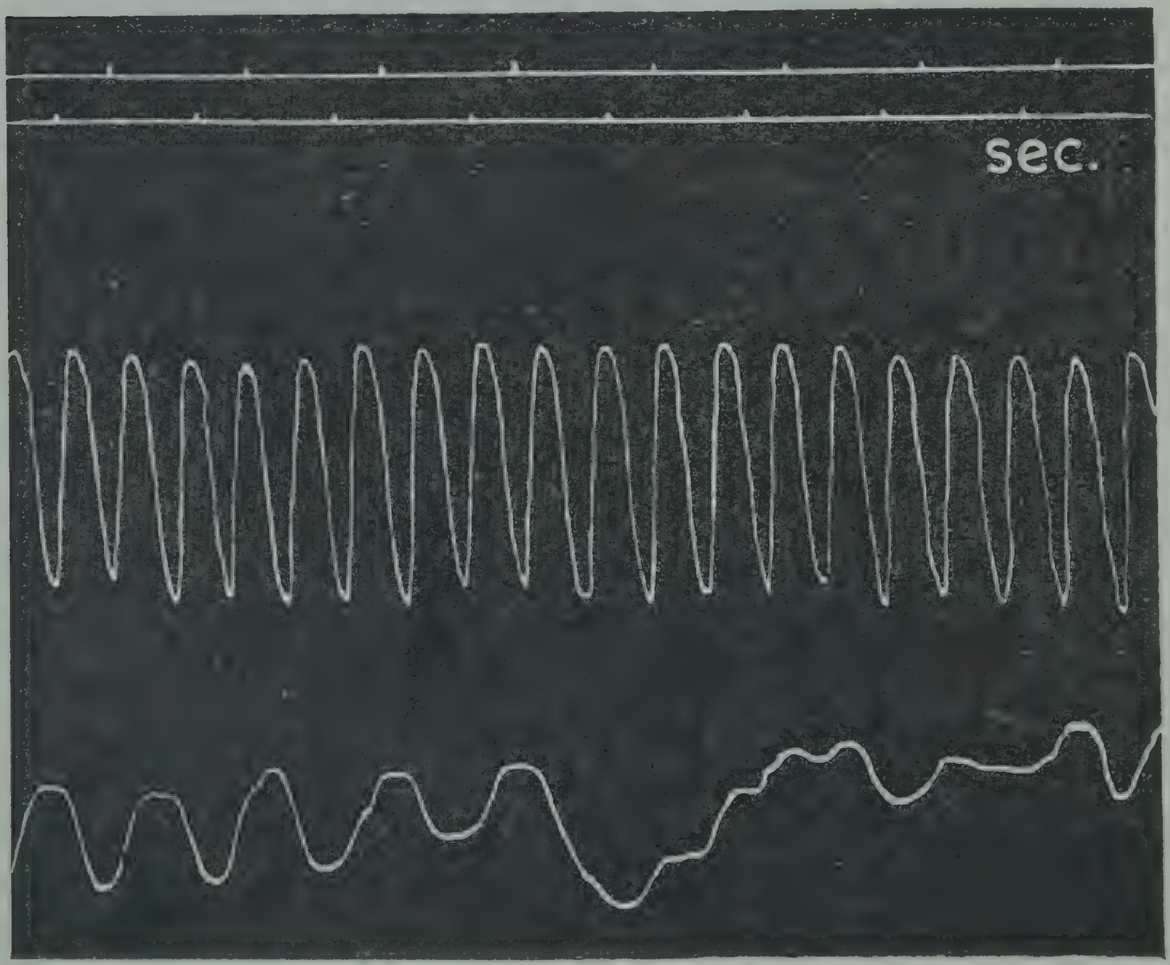


Fig. 47, 13

Tracings of rapid pronation-supination. The movements of the affected arm (below) were for a time regular though slower and of smaller amplitude than the normal, but grew irregular and the arm became more or less fixed in supination (Gordon Holmes (1939), *Brain*, **62**, 22).

affected side show the same kind of abnormal behaviour. The hypotonia in man is shown by an alteration in the knee jerk (Fig. 47, 10); since stretch reflexes depend on the tension in the muscles the knee jerk may be diminished in extent because of the hypotonia but the leg, if hanging free, tends to swing to and fro before coming to a stop (Fig. 47, 10, D). The most prominent disturbance of movement is intention tremor—any voluntary effort is accompanied by tremor, which may be very severe. On attempting to touch the nose with the finger the face may be struck violently for not only is the force of the movement excessive but it is so badly directed that the nose may not be touched at the end of the movement (*dysmetria*, Fig. 47, 12). On the other hand, willed movements may be carried out with an undue slowness and deliberation, a complex movement being *decomposed* into its several parts. Rapid movements of the limbs, such for example as rapid pronation and supination of the forearms, are faulty, the movements in the affected side being slower, irregular and jerky (*adiadochokinesis*, Fig. 47, 13). If a patient suffering from loss of cerebellar function is asked to look to the right the eyes are moved to the right quickly enough but they return slowly to the mid position only to be deviated quickly to the right once more. This nystagmus, like the defects of voluntary movement just described, is a defect of postural fixation; it is an intention tremor of the eye muscles. Speech in cerebellar lesions is slow and 'scanning,' each syllable in a word being pronounced separately as if it were a separate word. This can also be explained as an intention tremor of the muscles used in vocalization.

REFERENCES

Adrian, E. D. (1943). Afferent areas in the cerebellum connected with the limbs. *Brain*, **66**, 289–315.

Adrian, E. D. (1944). Localization in the cerebrum and cerebellum. *Brit. med. J.* ii, 137–140.

Fulton, J. F. (1949). *Functional Localization in the Frontal Lobes and Cerebellum.* Oxford: Clarendon Press.

Holmes, G. (1939). Cerebellum of man. *Brain*, **62**, 1–30.

Jansen, J. & Brodal, A. (1954). *Aspects of Cerebellar Anatomy.* Oslo: Johan Grundt Tanum.

Moruzzi, G. (1950). *Problems in Cerebellar Physiology.* Oxford: Blackwell.

Snider, R. S. & Stowell, A. (1944). Receiving areas of the tactile, auditory and visual systems in the cerebellum. *J. Neurophysiol.* **7**, 331–357.

CHAPTER 48

THE DIENCEPHALON

THE forebrain comprises the *diencephalon* consisting of the structures bounding the third ventricle and the *telencephalon* which includes the cerebral hemispheres and the basal nuclei. In the embryo the lateral walls of the caudal end of the forebrain become thickened and develop two large nuclear masses, termed the *thalami*. Caudal to each thalamus two elevations, the medial and lateral geniculate bodies, appear; these constitute the *metathalamus*. The ventral and inferior parts of the lateral walls and the floor of the third ventricle form the *hypothalamus*.

THE HYPOTHALAMUS

The hypothalamus consists of the posterior perforated substance, the corpora mamillaria, the tuber cinereum, the infundibulum and the neurohypophysis; other nuclei are found in the lateral wall of the third ventricle below the hypothalamic sulcus and are listed in Fig. 48, 1. All these structures are derived from the diencephalon. Various classifications of the nuclei of the hypothalamus have been made but for physiological purposes they can be divided into four groups: anterior (including the paraventricular and supra-optic nucleus), middle (including the tuberal, dorsomedial and ventromedial hypothalamic nuclei), posterior (including the posterior hypothalamic nuclei and the mamillary nuclei) and lateral. The supra-optic nucleus is noted for its extremely rich blood supply, there being 2500 capillaries per cu. mm. It has been claimed by Beattie and others that stimulation of the anterior part (anterior and middle nuclei) of the hypothalamus produces parasympathetic effects whereas stimulation of the posterior part (lateral and posterior nuclei) produces sympathetic effects, but this division of the hypothalamus into an anterior parasympathetic part and a posterior sympathetic part has not been universally accepted.

Although there is abundant physiological evidence that the hypothalamus receives impulses from a wide area and its discharges affect the viscera and muscles all over the body, yet it has been difficult to trace these pathways by anatomical methods.

The medial forebrain bundle, the fornix, and periventricular fibres from the thalamus all lead impulses into the hypothalamus. The fornix brings fibres from the hippocampus to the mamillary nuclei; it was formerly believed that these fibres were concerned in olfaction but it is very doubtful if the hippocampus has any olfactory function. From the cerebral cortex fibres pass to the medial nucleus of the thalamus and then to the hypothalamus. There are also direct fibre connexions between the frontal cortex and the hypothalamus. Fibres which ascend from the brain stem reticular formation to enter the hypothalamus seem to be concerned in maintaining alertness (see p. 908). The mamillary peduncles contain fibres from spinal nerves and probably from the ninth and tenth cranial nerves.

There are four important efferent pathways leaving the hypothalamus.

From the mamillary bodies there pass the mamillothalamic and mamillotegmental tracts. The mamillothalamic tract passes from the mamillary nucleus to the anterior thalamic nuclei from which impulses can be relayed to the gyrus cinguli. The dorsomedial nuclei of the thalamus are reached by periventricular fibres; this area of the thalamus projects in turn to the frontal cerebral cortex. The most important descending fibres arise in the lateral and posterior hypothalamic areas and pass to the reticular formation of the mid-brain, where they probably relay to the brain stem and spinal cord. The posterior lobe of the pituitary gland (neurohypophysis), derived from a pocket on the floor of the third ventricle, receives in man about 100,000 unmyelinated nerve fibres from the hypothalamus, chiefly in the supraoptico-hypophysial tract. Some of these fibres are said to reach the pars intermedia and a few may even reach the anterior lobe of the pituitary body but the activities of this region are probably entirely under hormonal control (see p. 1001).

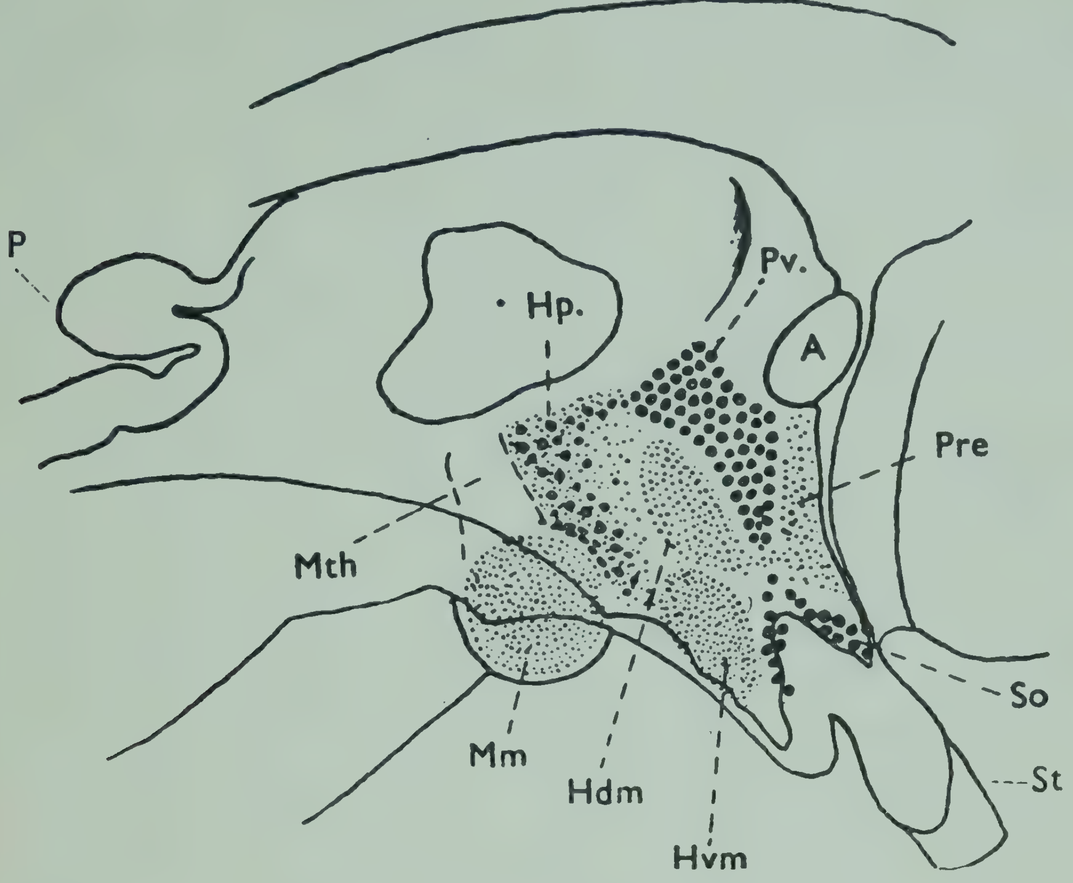


FIG. 48, 1

A diagram showing the hypothalamic nuclei projected on to the lateral wall of the third ventricle. This diagram has been made by taking a tracing from a photograph of the brain which was afterwards serially sectioned in a sagittal plane. The extent of the nuclei was reconstructed from the serial sections. × 2½ (W. E. Le Gros Clark (1936), *J. Anat.* **70,** 204).

Hp = nucleus hypothalamicus posterior.
Mth = mamillo-thalamic tract.
Mm = medial mamillary nucleus.
Hdm = nucleus hypothalamicus dorsomedialis.
Hvm = nucleus hypothalamicus ventromedialis.
So = nucleus supra-opticus.
Pre = nucleus pre-opticus.
A = anterior commissure.
Pv = nucleus paraventricularis.
P = pineal body.
St = stalk of hypophysis (infundibulum).

The hypothalamic sulcus runs from the aqueduct of the mid-brain along the lateral wall of the third ventricle as far as the interventricular foramen (near Pv in the diagram).

As mentioned earlier (p. 854), the decorticate animal with thalamus intact possesses normal muscular tone and shows righting reflexes. Although its internal environment is well regulated by autonomic reflexes, this is not so in the animal decerebrated by a section through the brain stem because the autonomic fibres from the posterior hypothalamus have been divided. When the cerebral hemispheres of the cat are removed without damage to the hypothalamus the animal shows, especially if the operation is recent, disturbances of behaviour, generally called 'sham rage', which appear to occur spontaneously or in response to a very mild stimulus such as a pinch. The animal spits, bites, lashes its tail, growls, and there is great activity of the sympathetic system as shown by erection of the hairs, dilatation of the pupils, salivation, increase of heart rate with extrasystoles and a rise of blood pressure. The hypothalamus can be stimulated through electrodes introduced into it under general anæsthesia. When the animal recovers consciousness stimulation through these electrodes, especially if they are placed in the lateral hypothalamus, produces, after a latent period of a few seconds, a similar sham rage. Animals deprived of the hypothalamic area but otherwise intact may show sham rage but never with the violence seen when this area is preserved.

The effects of electrical stimulation of various parts of the hypothalamus have been investigated by using small electrodes mounted in a stereotaxic apparatus which allows them to be placed accurately in any chosen position. Stimulation of the posterior hypothalamic areas produces sympathetic effects, such as rise of blood pressure, dilatation of the pupil, sweating, erection of hairs, inhibition of gut movements, and secretion of adrenaline from the suprarenal medulla. These sympathetic effects are best shown in unanæsthetized animals. In lightly anæsthetized animals, running or struggling movements may result from prolonged stimulation. The hypothalamus is not essential for sympathetic or parasympathetic effects since the cardiac reflexes controlling the heart rate function normally in an animal decerebrated through the pons; it is, however, called into play in conditions of stress. It seems likely that the hypothalamus influences the sympathetic system when the animal is exposed to cold (p. 205) and when there is emotional excitement. The description of these experiments shows that in addition to the visceral or autonomic component there is often a somatic component, such as a shivering of skeletal muscles, in the effects produced by the hypothalamus.

Stimulation of a region in the anterior part of the hypothalamic area causes a contraction of the urinary bladder, a parasympathetic effect. Since the effect persists after degeneration of fibres from the cerebral cortex, this region seems to be a true parasympathetic centre so far as the bladder is concerned. Although stimulation of the anterior part has produced other parasympathetic effects, such as increased gastro-intestinal activity and slowing of the heart, it may be that this is due to stimulation of fibres passing through the hypothalamic area from more cephalic parts of the brain.

The hypothalamus is essential for regulation of body temperature and the evidence for this has already been given (p. 205). Animals deprived of this part of the brain are poikilothermic. The main reactions to a cold environment, vasoconstriction and shivering, are abolished if the caudal area of the hypothalamus is destroyed; as in the case of the sham rage this reaction to cold involves both sympathetic and somatic (muscle) activity. The cranial area of the hypothalamus near the optic chiasma seems to be sensitive to a rise of body temperature and causes sweating and vasodilatation which lower body temperature. The descending pathways situated laterally in the caudal part of the hypothalamus carry impulses to affect the sympathetic system so as to cause cutaneous vasodilatation and sweating. Since these are purely sympathetic functions it is not justifiable to attribute parasympathetic activity to the anterior part of the hypothalamus. Because the hypothalamus is depressed by general anæsthetics, especially the barbiturates, deeply anæsthetized patients have no power to regulate body temperature.

All these observations fit in with the idea that the hypothalamus is the head ganglion of the autonomic nervous system but stimulation of the sensorimotor cortex and also of the temporal lobe produces alterations of blood pressure (either a rise or a fall) even when the hypothalamus is destroyed. It is not yet possible to say what is the physiological meaning of these recent findings.

The role of the pituitary in controlling the water balance of the body has already been described (p. 613). Removal of the posterior lobe alone produces only a transient polyuria but, if the stalk and median eminence are also removed, the diabetes insipidus is permanent. It is likely therefore that all these areas manufacture the antidiuretic hormone. Injection of acetylcholine or an anticholinesterase into the supraoptic nucleus releases antidiuretic hormone. Bilateral destruction of the supraoptic nuclei causes diabetes insipidus; the fibres of the supraoptico-hypophysial tract degenerate, the posterior lobe cells atrophy and production of antidiuretic hormone ceases.

Hypophysial tumours in man may be accompanied by obesity but this is due to damage to the middle group of hypothalamic nuclei. If hypophysectomy in animals involves damage to the hypothalamus, there is genital atrophy and obesity due to increase in appetite. The genital atrophy can be attributed to loss of the adenohypophysis. The relationship of the hypothalamus to ovulation and other reproductive functions is discussed on p. 970.

Injury to the hypothalamus, during surgical operations for example, may result in erosions or ulcers of the gastric mucosa, and is commonly followed by a persistent rise of body temperature; tumours of the posterior region may be associated with a fall of body temperature, a reduction in bodily activity (hypokinesia) and a tendency to fall asleep (hypersomnia).

THE THALAMUS

The thalami are two large masses of grey matter about 4 cm long placed on either side of the third ventricle and extending posteriorly, as the pulvinar, to overhang the superior quadrigeminal bodies. Each posseses three main nuclei; the anterior and medial form the *palæothalamus*, while the lateral nucleus, which occupies the greater portion of the thalamus including the pulvinar, constitutes the *neothalamus*. The geniculate bodies (the *metathalamus*) lie near the pulvinar, superolateral to the quadrigeminal bodies.

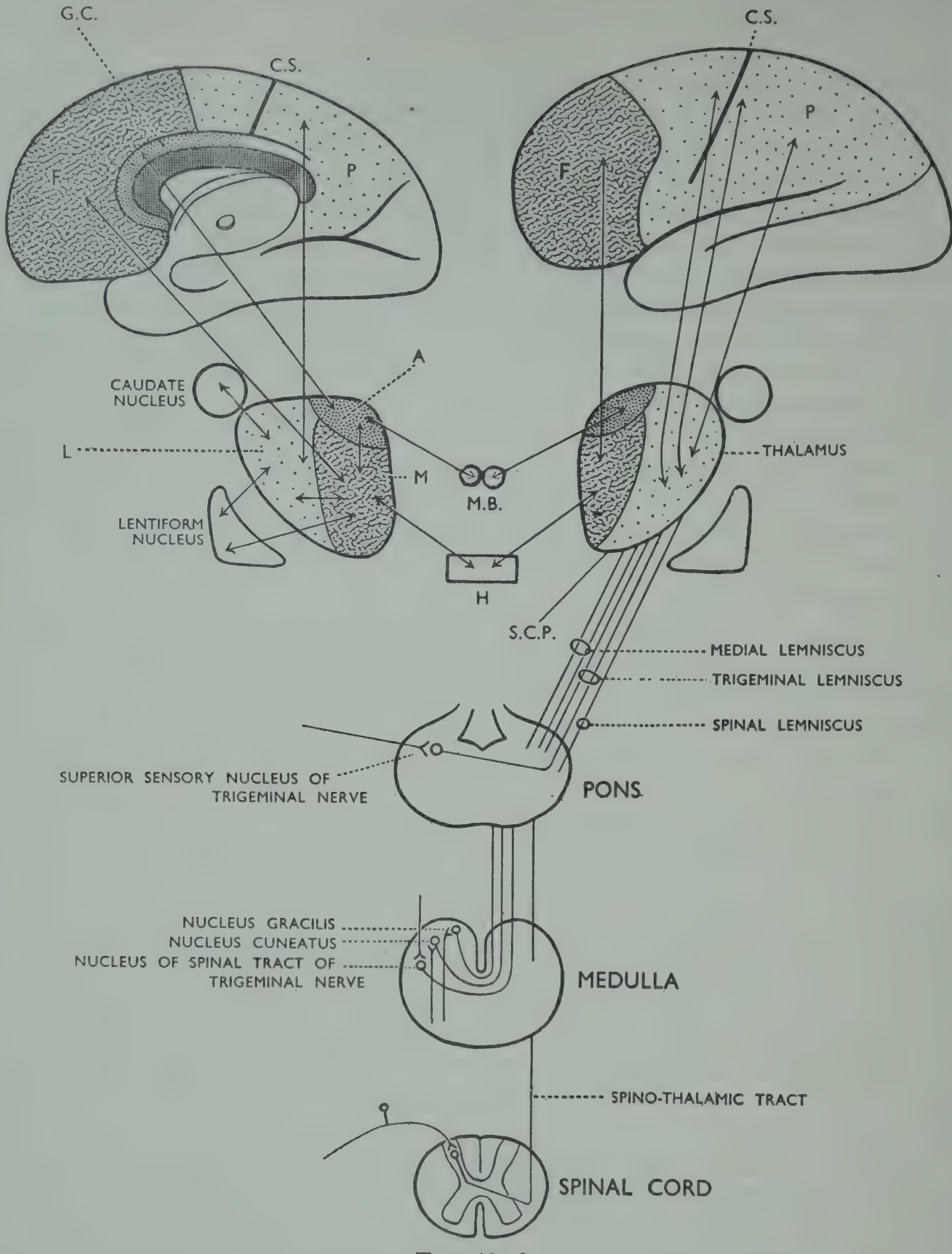

G.C.
C.S.
F
P
C.S.
F
P
A
CAUDATE NUCLEUS
L
M
M.B.
THALAMUS
LENTIFORM NUCLEUS
H
S.C.P.
MEDIAL LEMNISCUS
TRIGEMINAL LEMNISCUS
SPINAL LEMNISCUS
SUPERIOR SENSORY NUCLEUS OF TRIGEMINAL NERVE
PONS
NUCLEUS GRACILIS
NUCLEUS CUNEATUS
NUCLEUS OF SPINAL TRACT OF TRIGEMINAL NERVE
MEDULLA
SPINO-THALAMIC TRACT
SPINAL CORD

Fig. 48, 2

Cortical and spinal connexions of the thalamus.

G.C. = gyrus cinguli.
C.S. = central sulcus with motor and somesthetic areas on each side.
F = frontal area.
P = parietal area.
A = anterior group of thalamic nuclei.
M = medial group of thalamic nuclei.
L = lateral group of thalamic nuclei.
M.B. = mamillary bodies.
H = hypothalamus.
S.C.P. = superior cerebellar peduncle.

On the upper right of the diagram the projections are to the lateral surface of the cerebral hemisphere: on the upper left the projections to the medial surface of the cerebral hemisphere are given. Internuclear connexions of the thalamus are shown only on the left. The lines joining the various parts of the thalamus and cortex have in most cases been given arrows to indicate that impulses may pass in either direction. The metathalamus has been omitted. The pulvinar is included in the lateral nucleus.

(Courtesy of G. W. Pearce.)

The thalamus can be regarded as a relay station on the ascending sensory pathways (Fig. 48, 2). The anterior nucleus receives impulses from the mamillo-thalamic tract and transmits impulses to the gyrus cinguli. The medial nucleus receives some impulses via the hypothalamus from the viscera; most autonomic afferent fibres however relay in the lateral nucleus of the thalamus whence fibres pass to the sensory area of the cerebral cortex. Impulses can also pass from the frontal cortex to the medial nucleus and from there to the hypothalamus. The ventral portion of the lateral nucleus receives afferents from the medial, spinal and trigeminal lemnisci which carry impulses subserving proprioceptive, tactile, pain and temperature senses, and sends out efferents through the posterior limb of the internal capsule to the postcentral gyrus of the cerebral cortex. Thus the somatic sensory impulses originating in say the right side of the body cross over to the left thalamus which transmits them to the cerebral cortex on the left side only. There is physiological evidence that fibres from the surface of the body go to both thalami but the deep structures send fibres to the contralateral thalamus only. The lateral nucleus of the thalamus also receives fibres from the dentate nucleus of the cerebellum of the opposite side and relays impulses to the motor cortex (areas 4 and 6). Impulses from the cochlea pass through several synapses before reaching the medial geniculate body from which fresh impulses pass to the auditory cortex in the superior temporal gyrus (p. 902). The lateral geniculate body relays visual impulses from the retinæ to the calcarine cortex (area 17) on the same side (p. 900). These *cortical relay nuclei* do not act in a random way but relay impulses from points on a limb, for example, to specific areas on the cortex; similarly every point on the retina has a corresponding area on the calcarine cortex in spite of the intervention of the relay in the lateral geniculate body.

In addition to these cortical relay nuclei there are *association nuclei* in the neothalamus which receive impulses from the relay nuclei and project to the association areas of the cerebral cortex, namely the prefrontal, parietal and occipital cortex. The dorsomedial nucleus is the largest of these; it receives fibres from the lateral nucleus of the thalamus and from the hypothalamus and projects largely to the cortex of the frontal lobe. The thalamus also sends impulses to the caudate and lentiform nuclei, and the hypothalamus.

In cats, injection of minute quantities of strychnine into the thalamus increases the sensitivity of the skin to tactile, painful, or thermal stimuli on both sides of the body, but especially on the contralateral side. Since it is scarcely possible to get satisfactory information on sensation from animal experiments, reliance has to be placed chiefly on clinical studies. When the thalamogeniculate artery is blocked, the caudal portions of the thalamus are destroyed and there is a temporary loss of cutaneous sensation on the contralateral side of the body (a crossed hemianæsthesia) with permanent loss of sense of position of the limbs resulting in ataxia. After the lapse of a few weeks there may be intractable pain on the affected side. Stimuli, such as a pin-prick, which do not on the normal side give rise to anything more than discomfort, may cause quite severe pain. The threshold for pain is actually raised but the reaction to pain is exaggerated. There are no motor disturbances. This is described as the thalamic syndrome. The reason for these overactions is unknown; they do not occur when the sensory cortex alone is removed and cannot be due solely to a release from cortical control.

When the cerebral cortex is removed on one side (hemidecortication) in man there is for a few days anæsthesia on the opposite side, then nociceptive, that is painful, stimuli are appreciated but they are only poorly localized. Tactile sensation may eventually return but not the position sense. The return of crude sensation is partly due to the ability of the thalamus on the decorticate side to subserve crude sensations and partly to bilateral representation of touch and pain as mentioned above. Position sense is not bilaterally represented in the thalamus.

REFERENCES

Clark, W. E. Le G., Beattie, J., Riddoch, G. & Dott, N. M. (1938). *The Hypothalamus.* Edinburgh: Oliver & Boyd.

Fulton, J. F. & Ranson, S. W. (1940). *The Hypothalamus and Central Levels of Autonomic Function.* Baltimore: Williams & Wilkins Co.

Gaze, R. M. & Gordon, G. (1954). The representation of cutaneous sense in the thalamus of the cat and monkey. *Quart. J. exp. Physiol.* **39**, 279–304.

McDonald, D. A. (1951). W. R. Hess: the control of the autonomic nervous system by the hypothalamus. *Lancet,* i, 627–629.

Symposium (1940). The hypothalamus. *Res. Publ. Ass. Res. nerv. ment. Dis.* **20**. Baltimore: Williams & Wilkins Co.

Walker, A. E. (1938). *The Primate Thalamus.* Chicago: University Press.

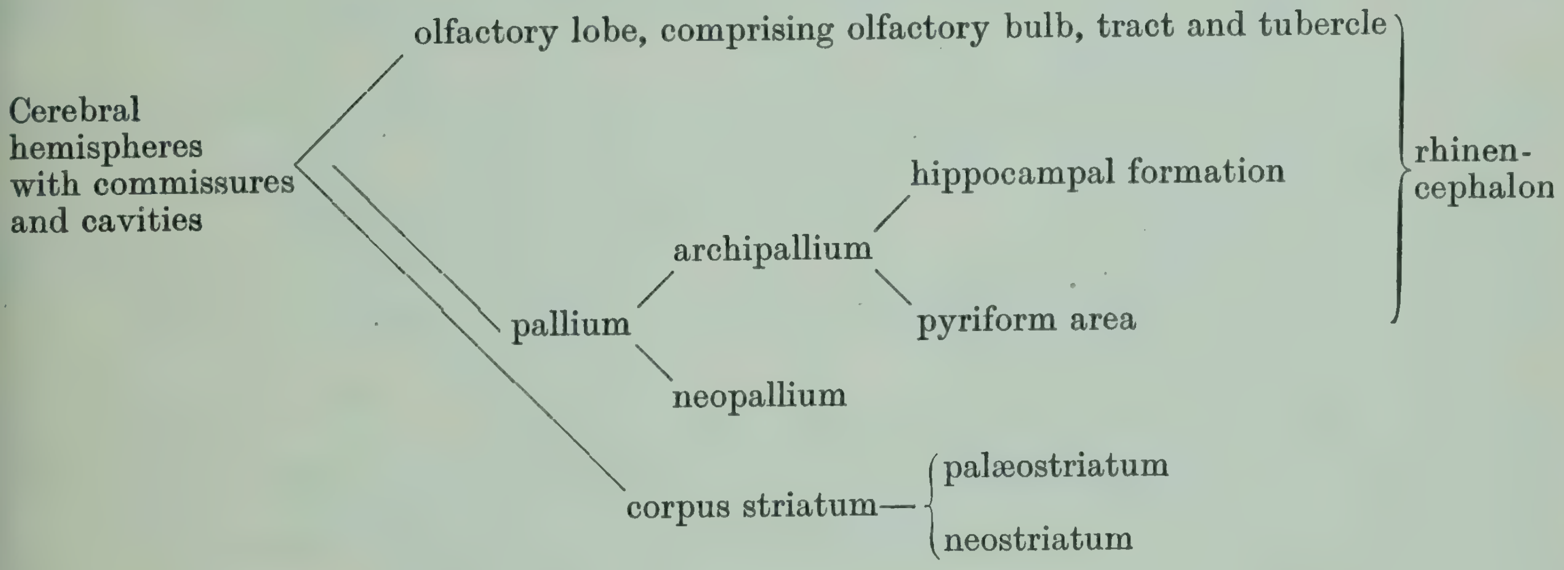
CHAPTER 49

THE TELENCEPHALON

THE telencephalon or end-brain is the most cranial (or rostral) part of the fore-brain and is represented in the adult by the cerebral hemispheres which consist of the olfactory system (collectively referred to as the rhinencephalon), the cortex or pallium and the corpus striatum. These divisions of the telencephalon are given in Table 49, 1. In the hemispheres are cavities called lateral ventricles each communicating with the central cavity, the third ventricle. The rhinencephalon is the part of the brain connected with the sense of smell and is found at the base of the brain. The pallium or cerebral cortex is divided into a small archipallium which is part of the rhinencephalon and a non-olfactory part, the neopallium. The corpus striatum consists of large nuclei of nerve cells situated within the substance of the hemispheres. The rhinencephalon, consisting of the archipallium and the olfactory lobe, is very small in man; it is much larger in lower animals in which the sense of smell is well developed. The piriform area is probably concerned with olfaction but the hippocampal area is unlikely to be. The remainder of the pallium, the neopallium, is the most highly developed part of the brain in man. It will be considered in detail in this chapter.

TABLE 49, 1

Parts of the Telencephalon


olfactory lobe, comprising olfactory bulb, tract and tubercle
Cerebral hemispheres with commissures and cavities
hippocampal formation
rhinen-cephalon
archipallium
pallium
pyriform area
neopallium
corpus striatum—
palæostriatum
neostriatum

The superficial layer or cortex of the hemispheres consists of grey matter containing numerous nerve cells, while the deeper part or white matter is made up of fibres connecting the various parts of the hemispheres with one another and with lower parts of the central nervous system such as the brain stem and spinal cord.

The surface of the cerebral cortex is marked by furrows called sulci, the areas between sulci being called gyri or convolutions. The two main sulci are the lateral and the central sulcus (Fig. 49, 2). The area in front of (rostral to) the central sulcus is the frontal lobe; the area immediately behind the central sulcus is the parietal lobe. The most (posterior) caudal part of the cerebral hemisphere is the occipital lobe. The area below the posterior ramus of the lateral sulcus is the temporal lobe. Smaller sulci divide up each lobe into numerous gyri, the names of the main areas being given in Fig. 49, 2.

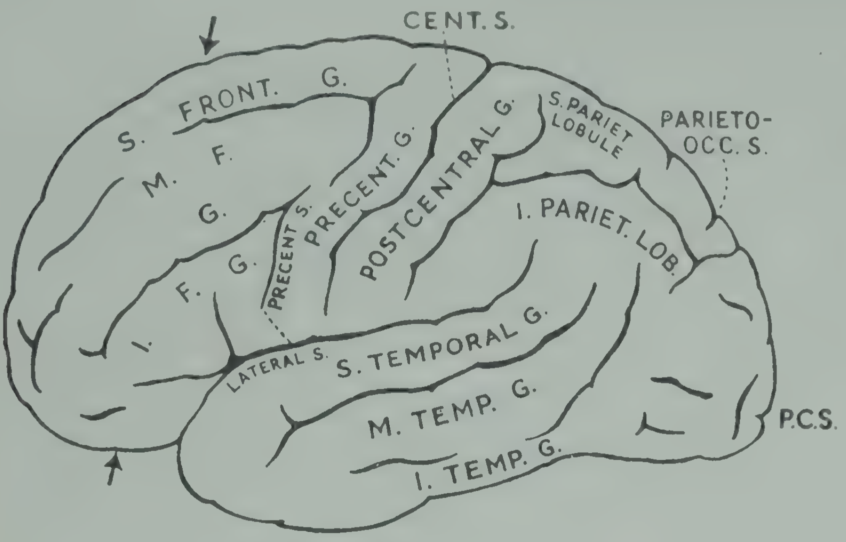

CENT. S.
S. FRONT. G.
M. F. G.
I. F. G.
PRECENT. S.
PRECENT. G.
POSTCENTRAL G.
S. PARIET LOBULE
PARIETO-OCC. S.
I. PARIET. LOB.
LATERAL S.
S. TEMPORAL G.
M. TEMP. G.
I. TEMP. G.
P.C.S.

Fig. 49, 2, A

The lateral aspect of the left cerebral hemisphere.

S. FRONT G. = superior frontal gyrus.
M. F. G. = middle frontal gyrus.
I. F. G. = inferior frontal gyrus.
PRECENT. S. = precentral sulcus.
PRECENT. G. = precentral gyrus.
POSTCENTRAL G. = postcentral gyrus.
S. TEMPORAL G. = superior temporal gyrus.
M. TEMP. G. = middle temporal gyrus.
I. TEMP. G. = inferior temporal gyrus.
S. PARIET. LOBULE = superior parietal lobule.
I. PARIET. LOB. = inferior parietal lobule.
CENT. S. = central sulcus.
PARIETO-OCC. S. = parieto-occipital sulcus.
P. C. S. = postcalcarine sulcus.
LATERAL S. = lateral sulcus (posterior ramus).

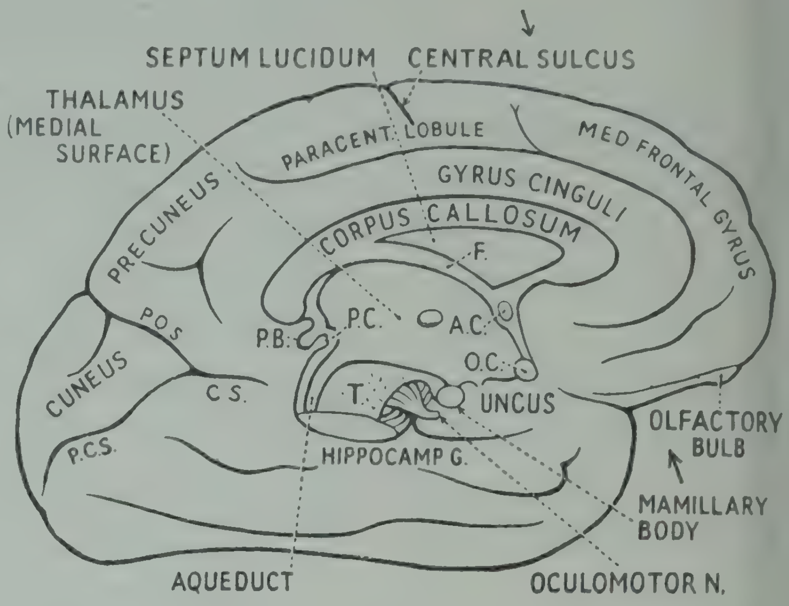

SEPTUM LUCIDUM
CENTRAL SULCUS
THALAMUS (MEDIAL SURFACE)
PARACENT. LOBULE
MED FRONTAL GYRUS
GYRUS CINGULI
CORPUS CALLOSUM
F.
PRECUNEUS
P.O.S
P.B.
P.C.
A.C.
O.C.
CUNEUS
C.S.
T.
UNCUS
OLFACTORY BULB
P.C.S.
HIPPOCAMP G.
MAMILLARY BODY
AQUEDUCT
OCULOMOTOR N.

Fig. 49, 2, B

The medial aspect of the left cerebral hemisphere.

A. C. = anterior commissure.
O. C. = optic chiasma.
P.-O. S. = parieto-occipital sulcus.
C. S. = calcarine sulcus.
MED. FRONTAL GYRUS = medial frontal gyrus.
P. C. S. = postcalcarine sulcus.
F. = fornix.
PARACENT. LOBULE = paracentral lobule.
HIPPOCAMP. G. = hippocampal gyrus.
P. B. = pineal body.
P. C. = posterior commissure.
T. = tegmentum.

A line joining the arrows gives the approximate plane of section in the operation of frontal leucotomy (p. 897).

BASAL NUCLEI

The basal nuclei (basal ganglia) include the corpus striatum (formed from palæostriatum and neostriatum), the amygdaloid nucleus and the claustrum. The corpus striatum is divided by the anterior limb of the internal capsule into the caudate nucleus medially and the lentiform nucleus laterally (Figs. 49, 3, and 49, 7), the latter being further subdivided into putamen and globus pallidus. The caudate nucleus and putamen contain small cells and together represent the neostriatum which receives many afferent fibres chiefly from areas 4 and 6 of the cortex and also from the thalamus. The globus pallidus or palæostriatum possesses large cells and may be regarded as the discharging centre of the basal nuclei. The efferent fibres leaving the palæostriatum make up a large part of the descending extrapyramidal system; they go to the thalamus, hypothalamus, subthalamic nucleus, red nucleus, substantia nigra and reticular substance of the mid-brain. From the functional point of view we can group together with the basal nuclei proper, the red nucleus, the substantia nigra and the subthalamic body of Luys although the last three are in fact situated in the brain stem.

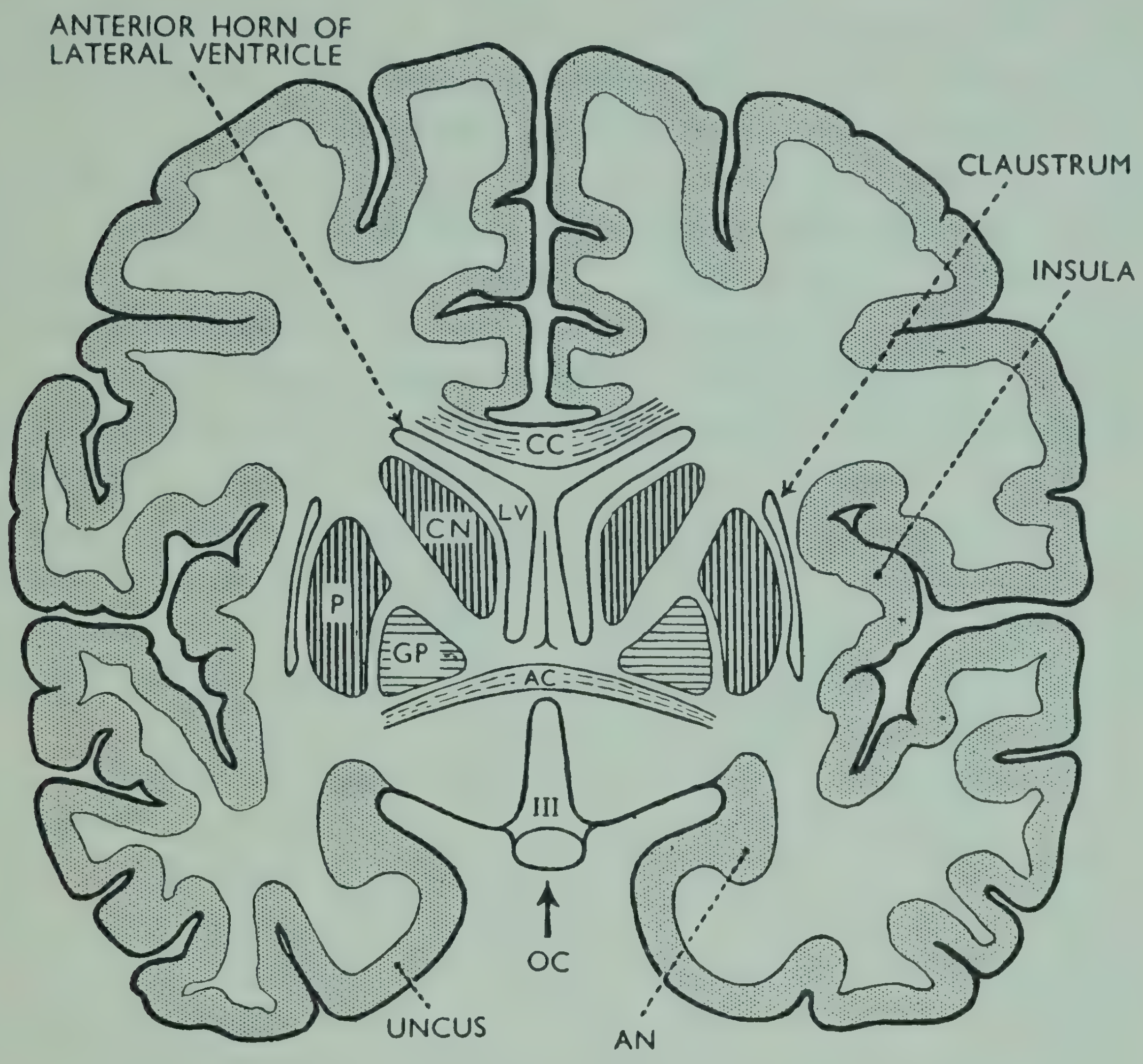


Fig. 49, 3,

Coronal section through the adult human brain just behind the optic chiasma.

▥ neostriatum (lentiform and caudate nuclei).

▤ palæostriatum (globus pallidus).

LV = lateral ventricle.
CN = caudate nucleus.
P = putamen.
GP = globus pallidus.
CC = corpus callosum.

OC = optic chiasma.
AN = amygdaloid nucleus.
F = anterior column of fornix
III = third ventricle.
AC = anterior commissure.

The basal nuclei, since they are quite prominent in primitive vertebrates, may be described as an old part of the fore-brain in a phylogenetic sense. They appear to have chiefly motor functions. Cats and dogs deprived of the cerebral cortex can carry out quite complex motor activities, like walking and running, but do not appear to be able to recognize their attendants or other animals. They eat only if food is placed in their mouths. If the basal nuclei are removed from such decorticate animals their behaviour is not greatly altered. It seems probable, therefore, that the basal nuclei are activated by the cortex almost entirely and that when the cortex is removed they cease to function. Destructive lesions of the basal nuclei in cats or dogs do not produce any visible effects unless neighbouring areas such as the internal capsule are also damaged. Although stimulation of the basal nuclei by electrodes inserted into them seldom has any apparent effect, it may inhibit a movement of the muscles on the opposite side which is being carried out as the result of cortical stimulation, and it is known that impulses from areas 4 and 6 of the cortex go to the caudate nucleus, then to the globus pallidus and thalamus and on to area 4 of the cortex. If this circuit is interrupted the restraining influence of the basal nuclei on muscle tone is lost and athetosis and other disorders of movement described below are seen.

The most common clinical disorder associated with damage to the globus pallidus and substantia nigra is *paralysis agitans* in which the patient shows a tremor when he is not making purposive movements; unlike cerebellar tremor (p. 870), it disappears during voluntary movement. All the muscles show a rigidity and in the face this results in an expressionless or mask-like appearance. The muscles show an intermittent (cog-wheel) resistance when they are moved passively. The arms are carried partly folded across the chest. The tremor disappears if such a patient develops a hemiplegia (paralysis of one side of the body) due to a hæmorrhage into the internal capsule, or if area 4 of the cortex (p. 890) is removed. Since in both cases the corticospinal (pyramidal) motor pathways are interrupted the tremor probably originates in the motor cortex.

Disease of the corpus striatum in man and injuries placed experimentally in this region in chimpanzees may produce involuntary behaviour called *athetosis* in which slow, twisting, sinuous movements of the limbs occur. Although these are associated with disease of the globus pallidus they may be due to compensatory extrapyramidal activity, since they are abolished by removal of the premotor area of the cortex (area 6). Damage to the caudate nucleus may result in quick, short, jerking movements, especially of the fingers, arm or face (*choreiform movements*). Lesions of the subthalamic body of Luys cause rhythmic, powerful, involuntary movements of a limb called *hemiballismus*.

CEREBRAL HEMISPHERES

Because the cerebral hemispheres are very much larger in man than in any of the primates, while the brain stem and spinal cord are not very different in size, we are led to conclude that human attributes are in some way associated with the greater development of the hemispheres. Such typically human activities as speaking and writing are extremely complex, involving as they do the co-ordination of sensory information from the eyes, the ears and various proprioceptors to produce a very delicately adjusted motor response. This implies the participation of many parts of the cortex and suggests that the sensory and motor areas must be interconnected by very numerous pathways. When these activities are analysed further it is found that they involve memory, learning and thought. There is no doubt that damage to the human cortex, especially if widespread, interferes with memory and judgment and we are, therefore, accustomed to think of these as cortical functions although no specific cortical area associated with memory has been discovered. Since the memory of an event, such as a pleasant meal, involves the recollection of visual, olfactory, gustatory and visceral impressions, it might be conjectured that memory is a function of the cortex as a whole rather than of any specialized part of it.

Although it is not to be expected that physiological studies will give much information about the more complex mechanisms of the brain, such as those of memory and thought, nevertheless the application of the methods used to explore the activity of the lower parts of the central nervous system has yielded information which has proved invaluable in the treatment of disease. The main clues to the understanding of cortical function in man are provided by the study of the effects of disease, which may either stimulate or destroy the affected areas of the cortex, and also of the effects of surgical removal and electrical stimulation, and by the investigation of its electrical activity. Histological examination of the cellular structure of the cortex and of the process of myelinization of the underlying white matter has also given important information.

HISTOLOGY OF THE CEREBRAL HEMISPHERES

In the fœtus at six months three different cellular layers can be recognized in the grey matter of the cortex and this arrangement persists in the adult archipallium. In the neopallium further differentiation takes place, as shown in Table 49, 4, and although six layers can be made out in the adult (Fig. 49, 6) the development varies in different parts of the

TABLE 49, 4

Structure of Cerebral Cortex

Six-months fœtus	Adult	Number of layer in Fig. 49, 6
Superficial or molecular layer of tangential fibres	Superficial or molecular layer of tangential fibres	1
Granular layer	Outer granular layer (small cells)	2
	Pyramidal layer (large cells)	3
	Inner granular layer (small cells)	4
Infragranular layer	Ganglionic layer (large cells in motor cortex)	5
	Polymorphous layer	6

cerebral hemispheres. The study of such histological differences, called architectonics, has received much attention during the last eighty years, and the cortex has been mapped out into areas which are numbered according to the system introduced by Brodmann (Fig. 49, 5). Area 4 of his classification, the *motor cortex*, is distinguished histologically by the large pyramidal cells and giant cells of Betz which may be 30 to 120 μ in height (layer 5, Fig. 49, 6). The *visuosensory cortex* (area striata or postcalcarine cortex) can be distinguished by the naked eye because the grey matter is divided by a white band due to an increase in the number of fibres in the outer band of Baillarger (Fig. 49, 6). The thickness of the grey matter of the cerebral cortex in man varies from about 4 mm to 1·25 mm. It is thinner in the occipital region than in the anterior and posterior central gyri.

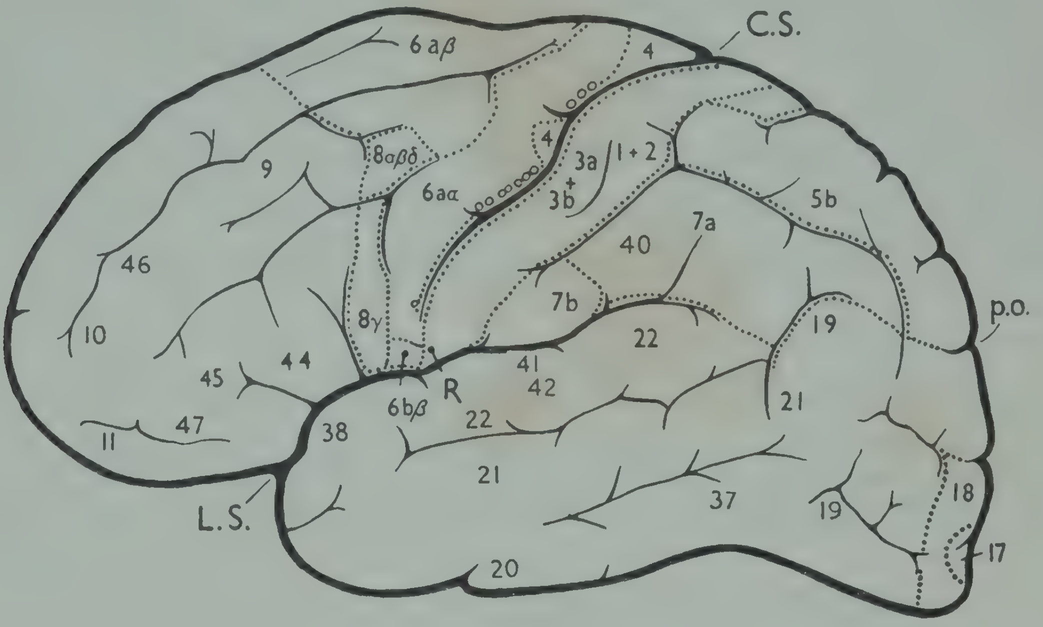
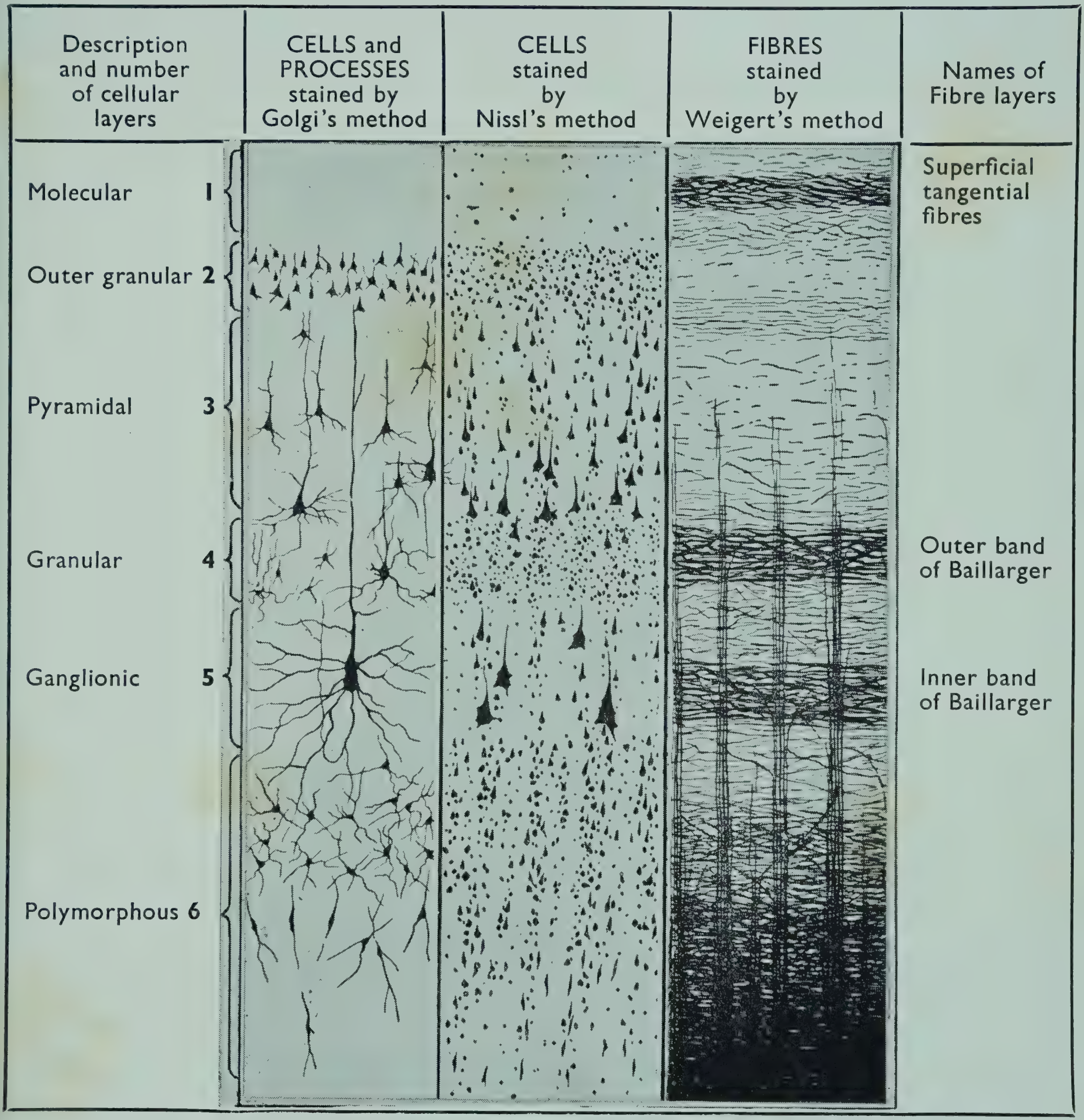


Fig. 49, 5

Architectonic fields of the lateral aspect of the human cortex according to Brodmann's conventional numbering. C.S. = central sulcus; L.S. = lateral sulcus; p.o. = parieto-occipital fissure; R = respiration; 4 = true pyramidal cortex. Stimulation of areas 5, 6, 7, also 8, produces adversive movements if the stimulation is strong enough. Stimulation of 8, 18, and 19 produces adversive eye movements. Stimulation of 22 produces eye and ear adversive movements. Stimulation of 6b produces mastication. Some of these fields extend on to the medial aspect. The gyrus cinguli is numbered 24 rostrally and 23 caudad. The names of the main gyri are given in Fig. 49, 2. (*After* W. Penfield & E. Boldrey (1937), *Brain*, **60**, 393.)

The white matter of the cerebral hemispheres consists of myelinated fibres which may run (1) as *association fibres* from one part to another part of the same hemisphere, or (2) as *commissural fibres* from one hemisphere to the other, or (3) as afferent or efferent *projection fibres* which carry impulses from or to the lower parts of the brain and the spinal cord. The neopallial parts of the cortex are joined by commissural fibres which pass across the mid-line in the corpus callosum (Fig. 49, 3). The main efferent neopallial projection fibres pass fanwise as the corona radiata, from all parts of the cortex, to run in the internal capsule which lies medial to the lentiform nucleus (Fig. 49, 7).

Nerve impulses have been shown by electrical methods to pass across the corpus callosum from one hemisphere to the corresponding area on the



Description and number of cellular layers
CELLS and PROCESSES stained by Golgi's method
CELLS stained by Nissl's method
FIBRES stained by Weigert's method
Names of Fibre layers
Molecular 1
Outer granular 2
Pyramidal 3
Granular 4
Ganglionic 5
Polymorphous 6
Superficial tangential fibres
Outer band of Baillarger
Inner band of Baillarger

FIG. 49, 6

Diagram to show the layers of cells and fibres in the grey matter of the human cerebral cortex (L. Luciani (1915), *Human Physiology*, **3**, 539. London: Macmillan).

other. It is rather surprising therefore to discover that this large tract of fibres can be cut through in man without producing any observable effect.

The fibres of the central nervous system acquire their myelin sheaths at different stages of development. The process of myelination begins at the sixth month of intrauterine life in the spinal cord and is seen at the eighth month in the sensory pathways going to the postcentral gyrus. Soon after this the afferent fibres of the visuosensory and auditosensory areas receive their myelin sheaths but the corticospinal (pyramidal) fibres are not completely myelinated until the child is three months old; the process continues in the psychic and association areas until the eighteenth year. Since it has been shown that a nerve fibre can conduct impulses only relatively slowly until it is myelinated, we have a physical basis for the continuous process of mental development up to adolescence. Because they are not yet myelinated it is unlikely that in the new-born child the cerebrospinal motor fibres are functioning normally and it may be that his motor activities are carried out chiefly through the brain stem and spinal cord.

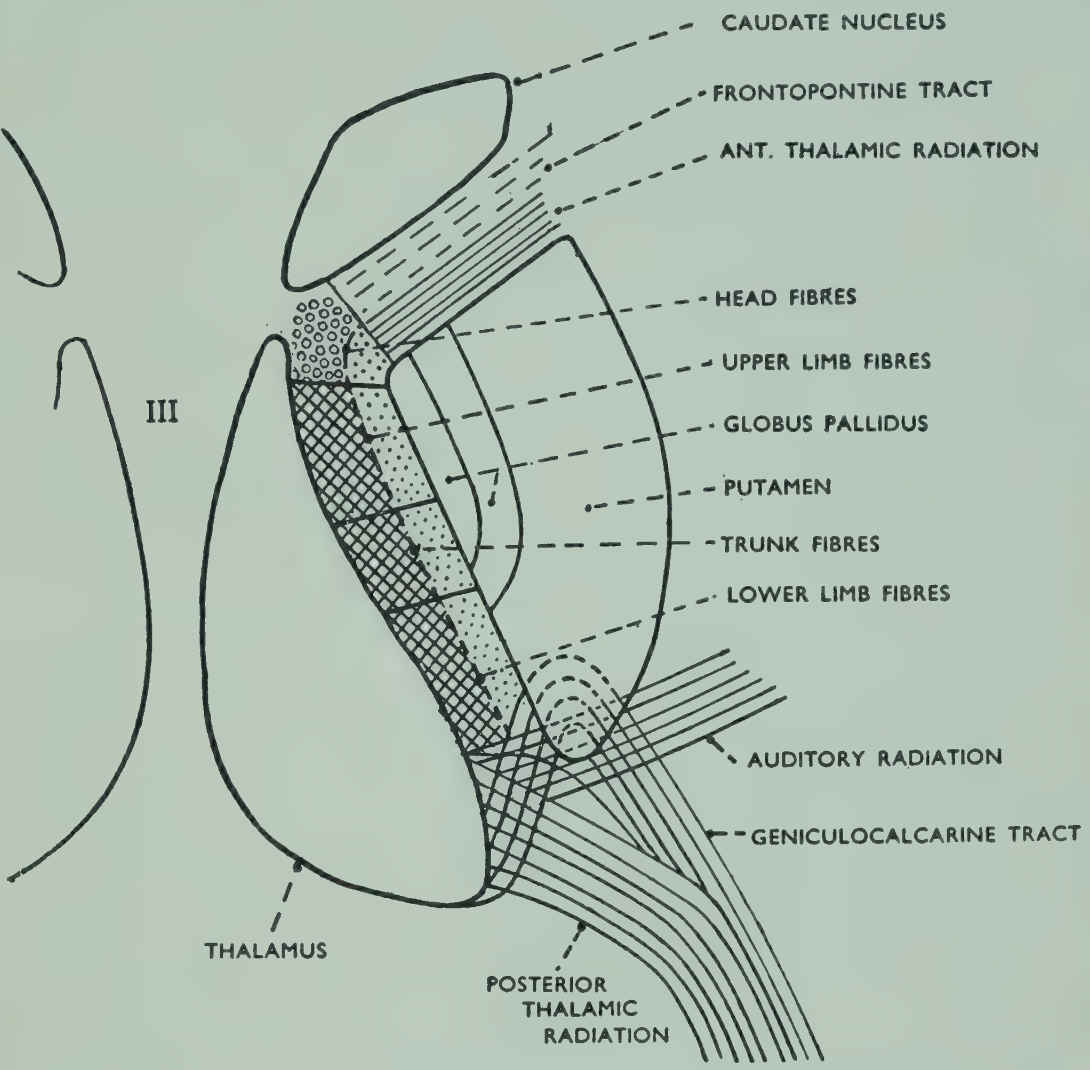

CAUDATE NUCLEUS
FRONTOPONTINE TRACT
ANT. THALAMIC RADIATION
HEAD FIBRES
UPPER LIMB FIBRES
GLOBUS PALLIDUS
PUTAMEN
TRUNK FIBRES
LOWER LIMB FIBRES
AUDITORY RADIATION
GENICULOCALCARINE TRACT
III
THALAMUS
POSTERIOR THALAMIC RADIATION

FIG. 49, 7

Diagram of the internal capsule. The sensory fibres of any part of the body emerging from the thalamus ▩ are medial to the motor fibres ░ of the same part. There is, however, a considerable admixture of motor and sensory fibres. (*After* S. W. Ranson & S. L. Clark (1951). *The Anatomy of the Nervous System.* 9th Edition p. 295. London: Saunders.)

MOTOR AREAS OF THE CEREBRAL CORTEX

In 1870 Fritsch and Hitzig found that electrical stimulation of the frontal cortex of the dog caused movements of the opposite side of the body. Ferrier in this country confirmed this a few years later in the monkey and showed further that removal of the area which on stimulation gave rise to movements of the hand resulted in paralysis of the hand. Numerous experiments in animals and man show that the area with the lowest threshold for electrical stimulation is area 4 of Brodmann (Fig. 49, 5) and this area is called, therefore, the *motor area* although there are, in fact, several other cortical areas concerned with motor function. This most easily excited region of the cortex contains the giant Betz cells and lies in man in that part of the precentral gyrus which is buried within the lips of the central sulcus. The following observations show that these cells are in fact important constituents of the motor area. In the first month of life before they are fully differentiated the electrical excitability of the area is low. By applying a hot plate for a short time to area 4 in the adult animal the superficial cells of the cortex can be destroyed while the Betz cells remain undamaged; it is then found that the excitability and spontaneous electrical activity of the area are practically unchanged. If the pyramids are sectioned in the medulla the Betz cells show chromatolysis.

The path of the axons of the Betz cells is shown by the degeneration which follows removal of area 4. The fibres pass down in the genu and in the anterior two-thirds of the posterior limb of the internal capsule in such a way that the fibres to the head are in front and those to the legs are posterior (Fig. 49, 7). These fibres then pass through the centre of the basis pedunculi and through the pons to the pyramids, where about 80 per cent. of the fibres cross to the opposite side in the decussation of the pyramids (Fig. 46, 2). There are about a million fibres in each pyramid but only about 34,000 Betz cells in each hemisphere, sufficient to account for only three per cent. of all the pyramidal fibres. The origin of many of the remainder is unknown, but to produce total degeneration of the pyramidal tracts it is necessary to destroy the whole hemisphere. The majority of the fibres in the pyramidal tract, about one-third of which are non-myelinated, are quite small in diameter; only 1·7 per cent. are in the range 10 to 22 μ and these may arise from the Betz cells. The vast majority of the fibres conduct relatively slowly. After removal of the motor area examination of the spinal cord shows degenerated synaptic knobs on the interneurones (p. 827) and not on the anterior horn cells. The corticospinal fibres, therefore, end at a position where the impulses carried by them may interact with impulses arriving by other descending tracts or through the posterior roots; the result of this interaction probably determines the nature of the impulses in the interneurones which go to the motor cells of the anterior horns.

In all higher mammals, including man, the part of the motor area nearest the vertex of the skull is concerned with movements of the leg, the middle part with those of the arm, while the lowest part of area 4, near the lateral sulcus, is concerned with movements of the face. The area of the cortical representation of any part is in proportion to the complexity of movements and not to the size of the part. Thus cortical representation of trunk movements is quite small, but that of the thumb, fingers, lips and tongue which are involved in skilled movements is relatively very large. In Fig. 49, 8, an attempt is made to illustrate this diagrammatically. Area 4 should not be regarded as a fixed mosaic of excitable cells with definite and circumscribed connexions with the periphery but as something with wide connexions, a certain degree of lability and a reserve of function as the following findings prove. Removal of a *small* part of the motor cortex does not result in permanent paralysis, the adjacent unharmed cortex eventually taking over the function of the ablated portion. The cortex is tolerant of destroying lesions and if such a lesion is restricted it may be symptomless. The connexions of any one area of the motor cortex are much wider than would be expected from the fixed mosaic idea. Thus after excision of a small part corresponding to the hand degenerated

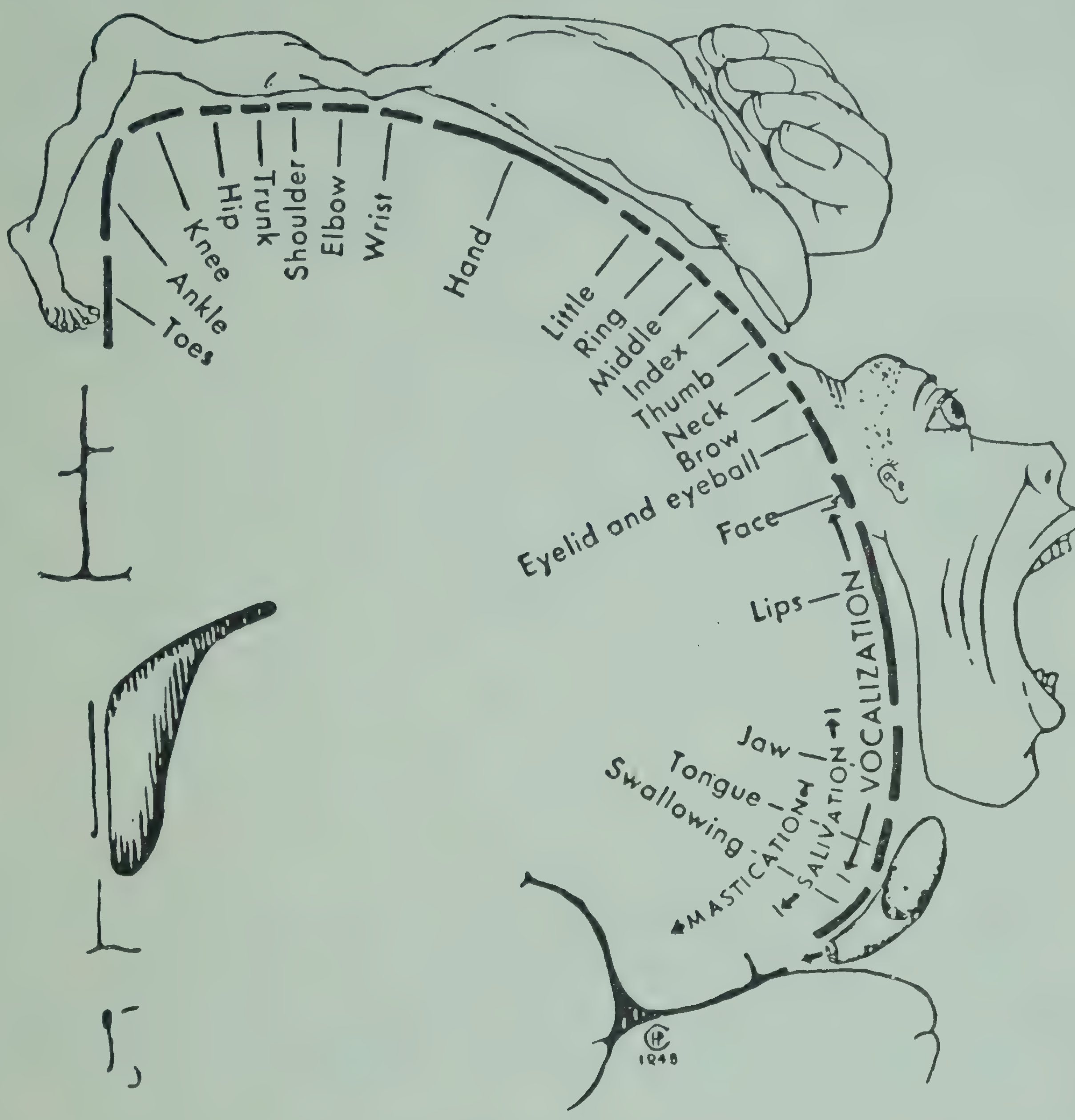

Toes
Ankle
Knee
Hip
Trunk
Shoulder
Elbow
Wrist
Hand
Little
Ring
Middle
Index
Thumb
Neck
Brow
Eyelid and eyeball
Face
Lips
VOCALIZATION
Jaw
Tongue
Swallowing
MASTICATION
SALIVATION
1948

FIG. 49, 8

Motor homunculus. The right side of the figurine is laid upon a cross-section of the hemisphere. This diagram gives the results of many experiments carried out at operation under local anæsthesia of the scalp. The exposed cortex was stimulated directly, usually by laying bipolar electrodes on it, the points of the electrodes being 3 mm apart. The stimulator provided 60 pulses of current per second. The potential required to elicit a movement varied between ½ and 5 volts. Stimulation of the cortex over the areas marked in thick lines produced movements of the muscles in the parts labelled (W. Penfield & T. Rasmussen (1950), *The Cerebral Cortex of Man*, p. 57. New York: Macmillan).

fibres can be traced to all parts of the spinal cord as far down as the lumbar region. It is well known that repeated stimulation of a single point on the cortex does not invariably lead to the same muscular movement. Finally, as described below, the actual map obtained varies with the method of stimulation.

In 1873 Hughlings Jackson described the initial signs of an epileptic fit due to a discharging lesion in the cortex in the following words. "There are three parts where fits . . . mostly begin, (1) in the hand, (2) in the face or tongue or both, (3) in the foot. . . . When the fit begins in the hand the index finger and thumb are usually the digits first seized: when in the face the side of the cheek is first in spasm: when in the foot almost invariably the great toe." It is difficult to believe that the cortical areas for the thumb, angle of the mouth and great toe (as shown on a classical map such as Fig. 49, 8) have a specially high incidence of discharging lesions. In fact such lesions may occur at any part of the motor cortex. An explanation of this apparent anomaly is afforded by the recent experiments of Liddell and Phillips illustrated in Fig. 49, 9. They used single rectangular pulses of current of about 5 msec duration to stimulate the exposed motor cortex of the baboon. The area representing the movements of the opposite thumb and index varied according to the strength of the pulses (Fig. 49, 9 (*a*)); the threshold was always lowest for the thumb complex. The areas for the thumb, hallux and mouth overlap as shown in Fig. 49, 9 (*c*) and when the strength of the pulse is sufficient movements in all three areas are obtained. If instead of single pulses a series of pulses at the rate of 50 per second is used, the simple large scale map (Fig. 49, 9) of distal parts changes into the traditional map of finer grain in which movements of all parts of the body are represented (Fig. 49, 8 was obtained in this way). The presence of these fields of low threshold explains adequately the common sites of onset of Jacksonian epilepsy.

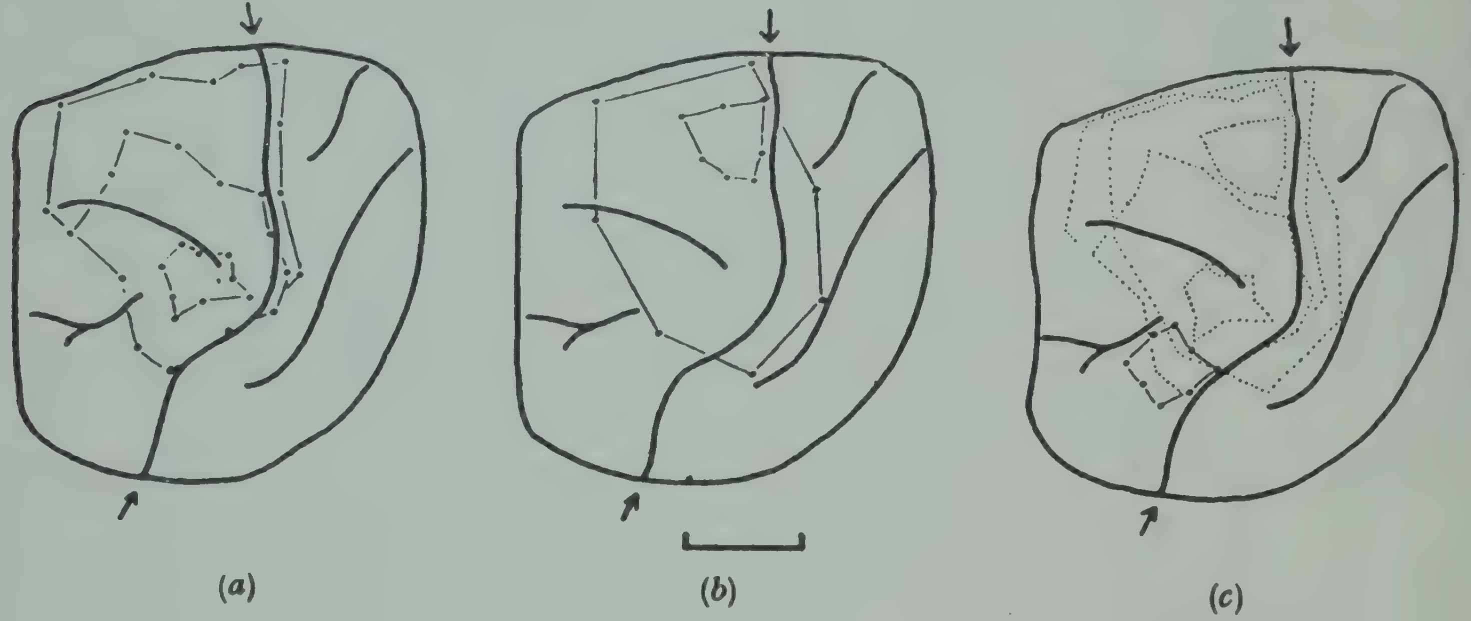


FIG. 49, 9

Female baboon (*Papio papio*). 5 kg. Left Rolandic region, stimulated with single pulses 4·5 msec in duration. Scale 10 mm. Arrows mark the central fissure. (Traced from the photographic map.) (*a*) Extent of thumb-index-minimus area at strengths 1·2 mA, 1·6 mA and 2·7 mA. (*b*) Extent of hallux-middle toe area at strengths 2·05 mA and 4·7 mA. (*c*) Area for angle of mouth (solid line) at strength 2·35 mA. Thumb and toe areas indicated by dotted lines (E. G. T. Liddell & C. G. Phillips (1950), *Brain*, **73,** 130).

If currents of suitable intensity and waveform are applied to area 4 (the motor area) in man, it is possible, according to some authors, to produce movements of single muscles on the opposite side of the body. Other investigators believe that area 4 deals with movements, not with individual muscles. In the words of Penfield and Rasmussen, "The movements produced by cortical stimulation are never skilled acquired movements but instead consist of either flexion or extension of one or more joints, movements which are not more complicated than those the newborn infant is able to perform." Area 6, the premotor area, immediately rostral to area 4, contains no Betz cells and is not so easily stimulated. Electrical stimulation of area 6 produces movements of small groups of muscles or more widespread movements involving perhaps a whole limb or a large part of the opposite side of the body (p. 890). If currents above threshold strength are applied to area 4, groups of muscles having the same action (synergic muscles) are activated but, if the current is made still stronger, movements begin in the same part of the body as before and spread to other parts of the body as adjacent motor cells are excited. The order of spread depends upon the anatomical arrangement of the excitable areas of the cortex; a movement starting in the thumb may spread to the fingers, forearm and shoulder, involving larger and larger groups of muscles, so that eventually the entire musculature on the opposite side of the body may be called into action. This is the sequence of events in a typical so-called Jacksonian fit, which may arise from irritation of the cortex as the result of damage to the skull or meninges. After such an epileptic fit, as after strong electrical stimulation, the excitability of the cortex declines or even disappears for a time, and the subject is inclined to fall asleep. This probably depends on phenomena already described in connexion with the positive after-potential (p. 745) and may be the result of a local fall in pH. Conversely a rise in pH, such as is produced by hyperventilation, increases the excitability of the motor area.

Epileptic seizures occurring spontaneously or by electrical stimulation of area 4 may occasionally involve the limbs on the same side. These ipsilateral seizures presumably depend on fibres coming down from the cortex but their route is at present unknown. In the lightly etherized macacus monkey stimulation of certain parts of the facial area of the motor cortex produces movements of the face on the same side whereas stimulation of adjacent areas gives contralateral or bilateral facial movements. There are other indications that in man the motor cortex is able to influence muscles of the same side. For example the recovery of motor function after unilateral ablation of the motor cortex may be due in part to the activity of fibres from the ipsilateral motor cortex. Even a man whose motor cortex has been removed on one side can learn to walk, that is, to use muscles on both sides of the body.

Effects of Removal of the Motor Cortex.—Removal of the whole cerebral cortex in dogs and cats has very little effect on walking or running but the hopping and placing reactions (p. 863) are lost. In the monkey, ablation of area 4 produces a *hemiplegia*, that is a paralysis or loss of voluntary movements on the opposite side of the body; immediately after this operation the muscles are hypotonic and the condition is described as a *flaccid paralysis*. Voluntary movement begins to return in a few days; first at the proximal joints and somewhat later in the fingers and toes. Although a remarkable degree of recovery occurs eventually, it is usually possible to detect a slight residual clumsiness in the fingers, for example in picking up small objects. In man after destruction of the motor cortex recovery is very much slower and dexterity of movements, especially the rapid and highly complex movements made with the fingers, is rarely regained.

If numerous intersecting cuts are made into the grey matter of the motor cortex of the monkey very little upset of motor function ensues. Horizontal, that is tangential, intracortical transmission cannot, therefore, be of any great importance.

When both pyramidal tracts are divided at the level of the medulla (Fig. 46, 4) in the monkey, the muscles become hypotonic so that the head sags and the abdomen protrudes. After some time, sitting and, later, walking become possible, but the animal is unwilling to move, and the ability to carry out accurate and delicate movements is lost. This experiment shows that a certain amount of voluntary movement can be carried out by extrapyramidal fibres since the loss of movement on section of the pyramidal tracts is very much less than that seen when area 4 and area 6 of the cortex are both removed. There is thus no doubt about the importance in locomotion of the extrapyramidal pathways.

As already mentioned, removal of area 4 in the primate causes a flaccid paralysis, but if the ablation extends forward from area 4 to area 6 a *spastic paralysis* results, just as is seen in man after hæmorrhage into the internal capsule. In man and other primates a spastic paralysis is characterized by an increase of flexor tone in the arms and of extensor tone in the legs. The hemiplegic patient, therefore, carries his paralysed arm flexed across his chest and his paralysed leg extended at the knee; there is loss of voluntary movement of the limb, but the limb is rigid and an attempt to flex it meets with resistance very like that observed in decerebrate rigidity; there is also exaggeration of the tendon reflexes and clonus can readily be demonstrated. Apart from the paralysis all these abnormalities are due to over-activity of the spinal stretch reflexes (p. 839). Since lesions restricted to the pyramidal tract in primates produce a hypotonic paralysis, the spastic paralysis seen after extensive cortical ablation or cerebral hæmorrhage in man must be presumed to be due to simultaneous interruption of pyramidal and extrapyramidal pathways.

When the corticospinal tract is interrupted by injury or disease the plantar reflex is extensor instead of flexor, that is to say the response produced by firmly stroking the sole of the foot from behind forwards produces dorsiflexion of the great toe (Babinski's sign). In addition, if the stimulus is strong enough, there may be flexion of the ankle, knee and hip, indicating that the reflex is actually part of a general flexor withdrawal reflex. Dorsiflexion of the great toe occurs also with destruction of cortical area 4 and in normal infants before the corticospinal fibres have become myelinated. If the cortical lesion involves the premotor area as well, stimulation of the sole of the foot produces not only dorsiflexion of the great toe but also a fanlike spreading of the other toes.

Hines (1937) has found that in the monkey removal of a narrow strip of cortical tissue, 4s, lying immediately rostral to area 4 produces spasticity with very little paralysis, and it has been suggested that spasticity occurs by the release of the lower parts of the nervous system from the inhibitory control exerted by this narrow area of cortex. It was later found that stimulation of the so-called 'suppressor' area 4s inhibits muscular tone and the muscular movements which follow stimulation of area 4. Later four additional suppressor areas were found in the cerebral cortex of monkeys, stimulation of which suppressor the activity of area 4. Then it was discovered that the application of strychnine to any of these suppressor areas caused a wave of suppression of electrical activity to pass slowly across the whole of the cerebral cortex. The axons of the cells of area 4s have been traced in the following way. After application of strychnine to a small area of the cerebral cortex, the neurones there discharge synchronously and a record of the electrical changes shows characteristic spikes which can be picked up in any region through which the axons of the strychninized neurones travel. In this way the majority of the axons from area 4s (probably also from most of the cortex) have been shown to travel along with the cerebrospinal fibres in the internal capsule, but to diverge from them in the brain stem and to pass finally into the inhibitory reticular formation in the brain stem (p. 850 and Fig. 46, 9); here the fibres relay by synapsing with the neurones of the reticulospinal fibres which pass down the lateral white columns of the spinal cord to end on the spinal interneurones. Stimulation of the ventro-medial medullary reticular area reduces the rigidity of decerebrate animals but the reticulospinal system does not act independently of cortical or cerebellar activation. The remainder of the fibres from area 4s travel to the caudate nucleus (p. 884), to the thalamus, and by thalamocortical fibres to area 4. Such cortico-cortical circuits have been found in all the suppressor areas and have been considered as the basis of the suppressive phenomena. Recent work has cast doubt on the notion of the existence of specific suppressor areas. For example, stimulation of 'suppressor' areas in the unanæsthetized monkey does not suppress motor activity and may even produce bilateral tonic movements. Further, under certain conditions of stimulation it has been possible to set up a slowly spreading wave of suppression from any area of the cortex of the monkey. Finally, the suppressor areas cannot be demonstrated in the unexposed cortex.

The reticular formation of the brain stem also contains fibres with a facilitatory effect, that is stimulation of these areas increases reflex effects such as the knee jerk. Facilitation of the spinal reflexes is also brought about by the pyramidal fibres. Spasticity may, therefore, be considered to be due either to loss of central inhibitory influences or to increase of facilitatory influence on the spinal stretch reflex, or to the operation of both factors simultaneously. The facilitatory fibres lie in the anterior half of the spinal cord, but as they are rather scattered any attempt to treat spasticity in man by making cuts in this area is liable to increase the incapacity.

Stimulation of the premotor area, for example area 6a*α*, usually causes movements of a small group of muscles—the same group that is activated by stimulation of the adjacent area 4. The effect depends on fibres running from area 6 to area 4 since after a cut has been made between the two areas, or after area 4 is removed, the character of the response is altered and stimulation of area 6a*α* yields only adversive movements consisting of rotation of the head and of the eyes to the opposite side with movements of the contralateral limbs. The impulses from area 6 must in these circumstances reach the spinal cord (probably by relaying in the reticular formation) by an extrapyramidal route. This is illustrated further by the finding that stimulation of area 6 produces movements of muscles even when the pyramids have been sectioned.

If a part of the premotor cortex, for example area 6a, is removed in the chimpanzee there is a slight weakness of the muscles of the opposite side but, although this soon disappears, skilled movements are permanently disturbed. This also applies to man after removal of the premotor cortex. A destructive lesion of the upper part of area 6 produces on the opposite side of the body a phenomenon called *forced grasping*. If the palmar surface of the hand or fingers is touched reflex flexion of the hand occurs to grasp the stimulating object and any attempt to withdraw the latter is followed by a firmer grasp. Since this reflex is normally present in infants its appearance in the adult is considered to be a regression to the infantile stage of the function of grasping. The grasp reflex is not seen after section of the pyramids alone, hence area 6 normally inhibits flexor activity through extrapyramidal pathways.

In addition to the main motor and premotor areas there are, as mentioned before, other areas of the cortex which are concerned in muscular movements. For example areas 44 and 45 together comprise the premotor area of the face (Broca's area); in right-handed persons lesions of this area on the left side produce motor aphasia, that is an inability to speak although the muscles of articulation are themselves not paralysed. Stimulation of the frontal eye fields in the posterior part of the second frontal convolution (area 8) in man produces conjugate deviation of the eyes to the opposite side without head movements and without visual hallucinations. Stimulation of the occipital eye fields (areas 18, 19) causes visual hallucinations and movements of the eyes to the opposite side. From area 17 visual hallucinations are obtained but not eye movements.

Cortical Representation of Autonomic Activities

The visceral or autonomic activities of the body seem to be controlled in the hypothalamus but there is a certain amount of cortical representation of autonomic activities, both sympathetic and parasympathetic. Stimulation of some points in areas 4 and 6 on the frontal lobe of the chimpanzee gives a rise of blood pressure and an increase of heart rate, but if the stimulating electrodes are moved only a short distance along the cortex the reverse effects may be obtained. Some of these effects may occur independently of the hypothalamus (p. 878). Injury to the cortex may produce vasomotor effects in man. Soon after a cerebral hæmorrhage which gives rise to a hemiplegia there is a cutaneous vasodilatation and an increase in sweating on the affected side; at the same time the surface temperature of the paralysed limbs is increased, although later it usually falls to subnormal levels. These autonomic effects are especially pronounced on stimulation of the orbital and cingulate gyri. In dogs stimulation of the cingulate gyrus arrests the movements of the pyloric antrum. Since destruction of the hypothalamus or section of the vagus abolishes this effect but section of the splanchnic nerves does not, the chain of connexions must be cerebral cortex, hypothalamus, medulla and vagus nerves to the stomach. Removal of the frontal lobes in monkeys increases the movements of the gastro-intestinal tract and the animals become

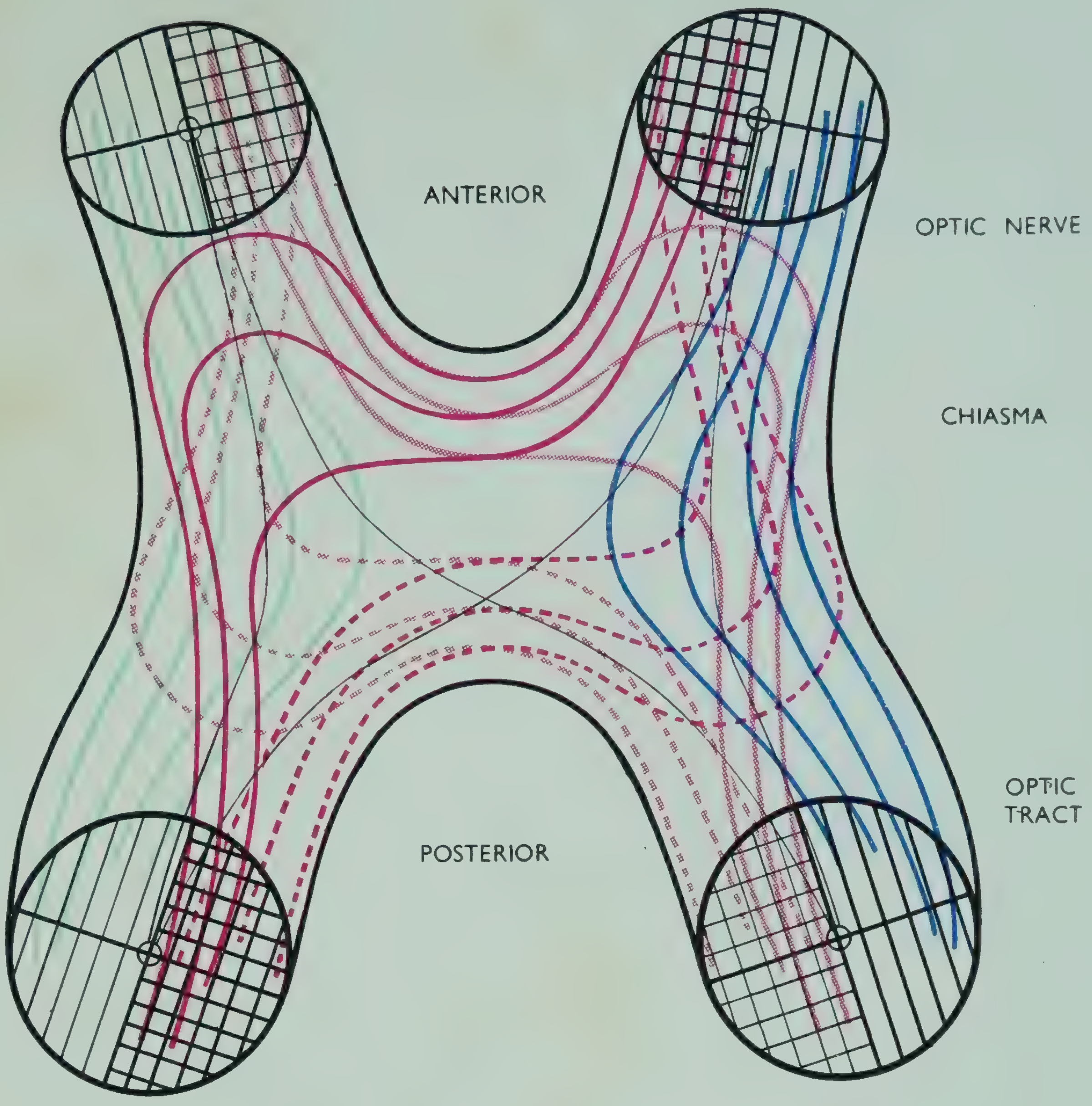


FIG. 49, 10

Diagram to show the probable arrangement of the chiasmal crossing in detail.

Continuous red lines indicate crossed fibres from the lower medial quadrants of the optic nerve; broken red lines crossed fibres from the upper medial quadrants; continuous blue lines uncrossed fibres from the lateral quadrants. In order to avoid confusion the fibres from the right optic nerve are coloured brightly, those from the left nerve faintly. The crossed fibres from the lower medial quadrant of the right optic nerve (continuous red line) are shown crossing in the anterior part of the chiasma and passing forwards into the termination of the left optic nerve, forming a knee there, and then backwards to the medial side of the left optic tract. The crossed fibres from the upper medial quadrant of the right optic nerve (broken red line) pass backwards along the same side of the chiasma to its posterior angle and then across its posterior part to the upper medial quadrant of the opposite tract. The uncrossed fibres from the lateral half of the nerve are spread out in the body of the chiasma so that those from the upper quadrant lie medial to those from the lower. At each side of the chiasma crossed fibres from both sides are mingled with uncrossed fibres of the same side. The position of the macular fibres is indicated by the thin black lines. These fibres occupy a relatively large area in the postero-inferior part of the chiasma and are therefore easily affected by interference from below and behind H. M. Traquair (1949), *Clinical Perimetry*, 6th ed., p. 78. London: Kimpton).

[*Facing page* **896**

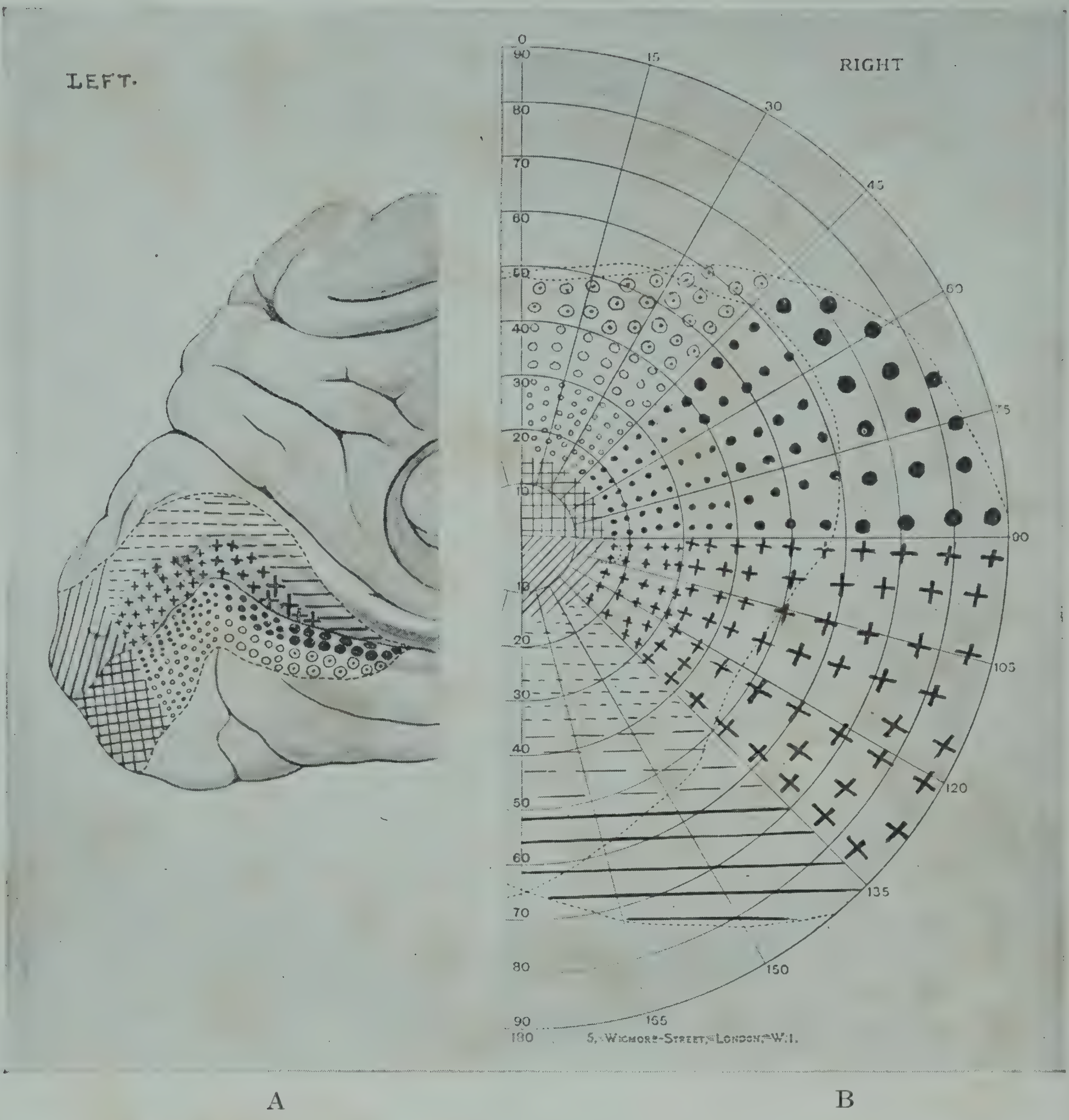

LEFT.
RIGHT
0
90
80
70
60
50
40
30
20
10
10
20
30
40
50
60
70
80
90
180
15
30
45
60
75
90
105
120
135
150
165
5, Wigmore-Street, London, W.1.
A
B

Fig. 49, 11

The cortical retina.

A diagram of the probable representation of the different portions of the visual fields in the calcarine cortex. No great accuracy is claimed for this schema

A. The postcalcarine fissure on the medial surface of the left occipital lobe is represented as widely opened. The macular area is relatively large, the peripheral area relatively small.

B. The right halves of the field of vision are superimposed.

Note that the centre of the field (maculæ) is represented at the tip of the occipital lobe and that the lower half of the field of vision (upper half of the retinæ) is represented in the cortex above the postcalcarine sulcus. The peripheral parts of the field of vision tend to be represented in the anterior portion of the calcarine area.

(*After* diagrams by Sir Gordon Holmes (1918), *Brit. J. Ophthal.* **2,** 383; also (1945), *Proc. Roy. Soc.* B, **132,** 348.)

exceedingly hungry; stimulation of area 6 may inhibit peristaltic activity. The secretion of saliva in response to a conditioned stimulus depends on the integrity of the lower part of area 6a. Stimulation of area 8 in the monkey produces a dilatation of the pupil whereas stimulation of area 19 produces a constriction. The pathways for these autonomic effects are known with certainty in only a few instances.

FRONTAL ASSOCIATION AREAS

The prefrontal area consists of that part of the frontal lobe rostral to areas 6 and 8. Electrical stimulation of this area causes no alteration of posture or of reflex behaviour. Removal of one prefrontal region produces no detectable changes in man or monkey. Removal of both prefrontal areas in monkeys results in restlessness and in defects in behaviour and learning which, however, only become obvious when quite complicated problems are presented to the animal. The normal chimpanzee given a complicated problem may be emotionally upset if it fails to solve it but after removal of both prefrontal areas this emotional response is not seen. In 1936 these findings were applied to man for the treatment of mental disorders by Moniz, a Portuguese neurologist, who devised the procedure of prefrontal leucotomy. The operation, performed through small holes made in the skull or in the roof of the orbit, divides the white matter of the prefrontal area (see Fig. 49, 2) and interrupts the pathways between the cortex and the thalamus and hypothalamus as well as connexions with other cortical areas. The main pathway interrupted is that to the large dorsomedial nucleus of the thalamus. Bilateral leucotomy has been found to benefit patients with mental illness, especially when 'psychic tension' is a prominent feature. If the section is made too far back, certain undesirable sequelæ such as labile emotional reactions, impaired moral sense and loss of initiative appear. Lesions restricted to the orbital or medial parts of the frontal lobes rarely show unpleasant after-effects.

SENSORY AREAS OF THE CEREBRAL CORTEX

It has been pointed out already in dealing with the special senses that objective investigation of sensation is possible only to a limited extent. Since little information can be gained from animal experiments, except perhaps by the study of conditioned reflexes, most of our knowledge must come from clinical investigation. For example, a conscious patient may describe his sensations when areas of the cortex exposed at operation are stimulated electrically. Data have also been obtained from the investigation of patients with traumatic or pathological lesions, although in such cases we are dependent on the co-operation of a patient who has a lesion the extent of which may be difficult to judge even at a later post-mortem examination. Information obtained in this way is not always entirely satisfactory.

The sensory areas of the cortex are the highest points in the brain reached by the impulses concerned in sensation. Although they are undoubtedly important as relay or junctional areas in the complicated pattern of the cerebral neurones, we are not justified in claiming that these areas are the seat of sensations and consciousness.

We must now consider the areas of the cortex concerned in vision, hearing, taste and touch. There is no cortical representation of pain.

OCCIPITAL CORTEX

Although it has been known for more than a hundred years that removal of the cerebral cortex produces blindness without interfering with the reaction of the pupils to light, only comparatively recently has the precise area of the human occipital cortex involved been localized to the walls of the postcalcarine sulcus on the medial surface of the occipital lobe (Brodmann's area 17), occasionally extending a short distance on to its lateral surface.

The structure of the retina has already been described and the axons of the ganglion cells have been followed into the optic nerve. The path of the fibres after this has been found by studying the degeneration occurring after removal of different parts of the retina. The fibres from the nasal half of each retina in man cross over at the optic chiasma (Fig. 49, 10, 12 and 13) so that the fibres from the right halves of the two retinæ pass into the right optic tract to the right lateral geniculate body from which new fibres go as the *optic radiations* to the calcarine cortex. The macular fibres concerned in central vision lie nearly centrally in the optic tracts and end on the posterior part of the lateral geniculate body, the macular relay fibres then going to the tip of the occipital lobe. Removal of the occipital cortex in man has often left some macular vision but this is likely to depend on incomplete removal of the area of cortex concerned rather than to bilateral cortical representation of the macula for which there is no anatomical basis. The results of correlation of the position of lesions with the shape of the visual field obtained by perimetry show that the centre of the field is represented in the tip of the occipital lobe at the posterior end of the postcalcarine sulcus, and that the upper half of each retina (lower half of the field of vision) is represented on the upper part of the visual area of the cortex, while the lower half of the retina is represented on the lower half. The area of the cortex connected with the macula and devoted to central vision is, as might be expected from the power of the macula to appreciate colour and fine detail, large compared with that subserving peripheral vision (Fig. 49, 11). Removal of both occipital lobes in man results in permanent blindness, but in dogs and rats recovery is sufficient after a time to allow differences in light intensity to be distinguished.

The areas adjacent to the occipital cortex, Brodmann's areas 18 and 19, are association areas in which co-ordination of eye reflexes with other reflexes occurs. The interpretation of what is seen, for example a printed word, requires the presence of these association regions.

Visual Reflexes.—When the superior quadrigeminal bodies are stimulated electrically dilatation of the pupil is said to occur along with conjugate movements of the eyes towards the opposite side, that is lateral movements of both eyes simultaneously through equal angles. It is possible however that these results, reported many years ago, are due simply to spread of current to neighbouring structures. When the superior quadrigeminal bodies are destroyed the pupillary contraction on exposure to light remains but, if the pretectal area just rostral to them (Fig. 46, 8) is destroyed at the same time, the light reflex can no longer be obtained. The explanation for this is clear from Fig. 49, 13; some fibres from the retina travel in the optic nerve and optic tract but pass medial to the lateral geniculate bodies to end in the pretectal area where the first synapse occurs. Second order relay neurones cross in the posterior commissure and reach the Edinger–Westphal part of the third nerve

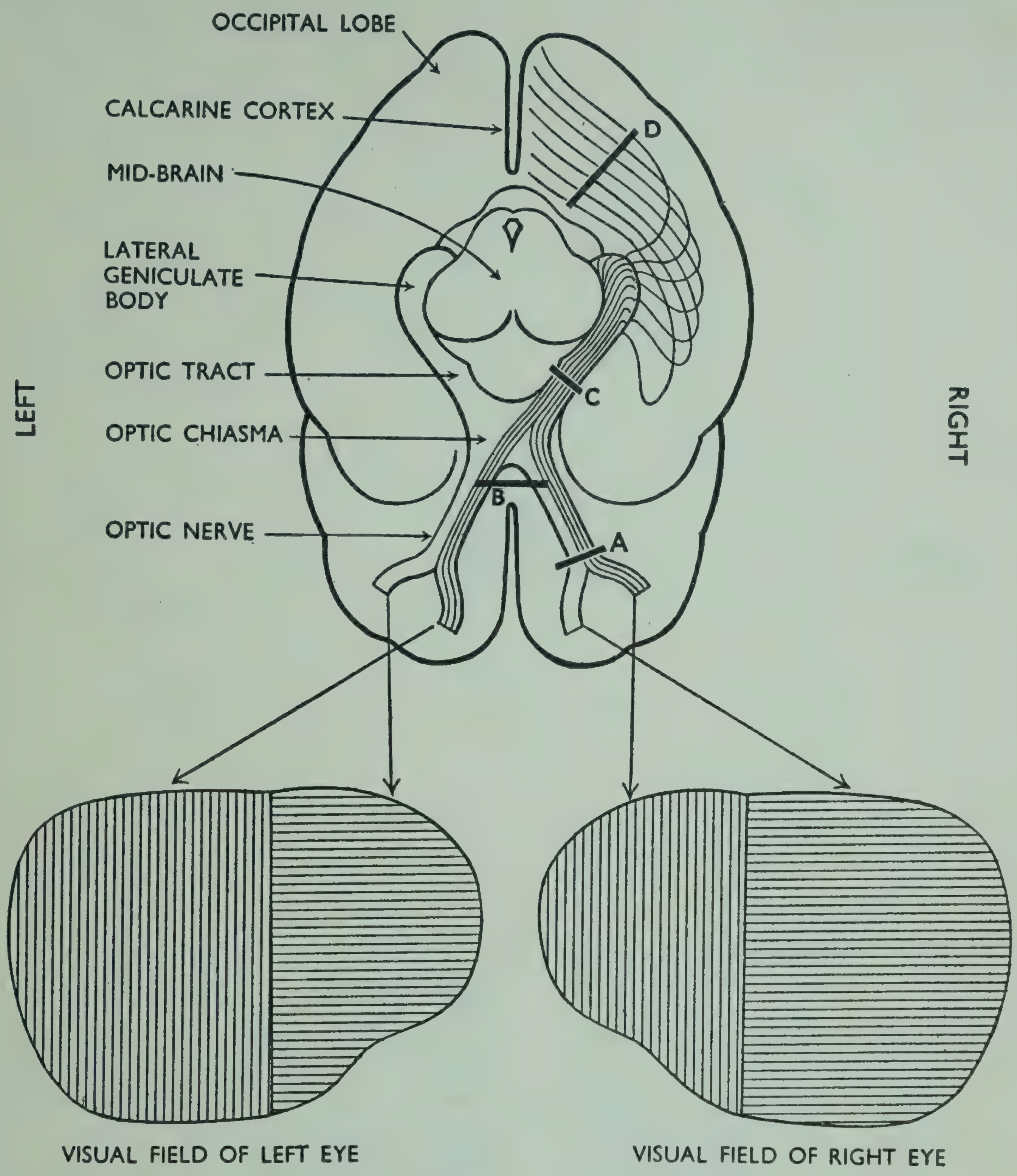

OCCIPITAL LOBE
CALCARINE CORTEX
MID-BRAIN
LATERAL GENICULATE BODY
OPTIC TRACT
OPTIC CHIASMA
OPTIC NERVE
LEFT
RIGHT
D
C
B
A
VISUAL FIELD OF LEFT EYE
VISUAL FIELD OF RIGHT EYE

FIG. 49, 12

The course of the fibres involved in vision from the right halves of the two retinæ to the occipital cortex. These correspond to the left halves of the fields of vision. There is a relay in the lateral geniculate body. A lesion at A cutting through the right optic nerve results in complete blindness of the right eye. A lesion at B involving the rostral part of the optic chiasma interrupts the fibres from the nasal halves of the retinæ and produces bitemporal hemianopia. A lesion at C divides the optic tract containing fibres from the right halves of both retinæ and produces blindness in the left halves of the visual fields (left hemianopia); the pupil does not contract when a light is thrown on the right halves of the retinæ. A lesion at D cutting through all the optic radiations also produces blindness in the right halves of the retinæ, but the light reflex is normal (Sir J. Conybeare (1947), *Textbook of Medicine,* 8th ed., p. 887. Edinburgh: Livingstone).

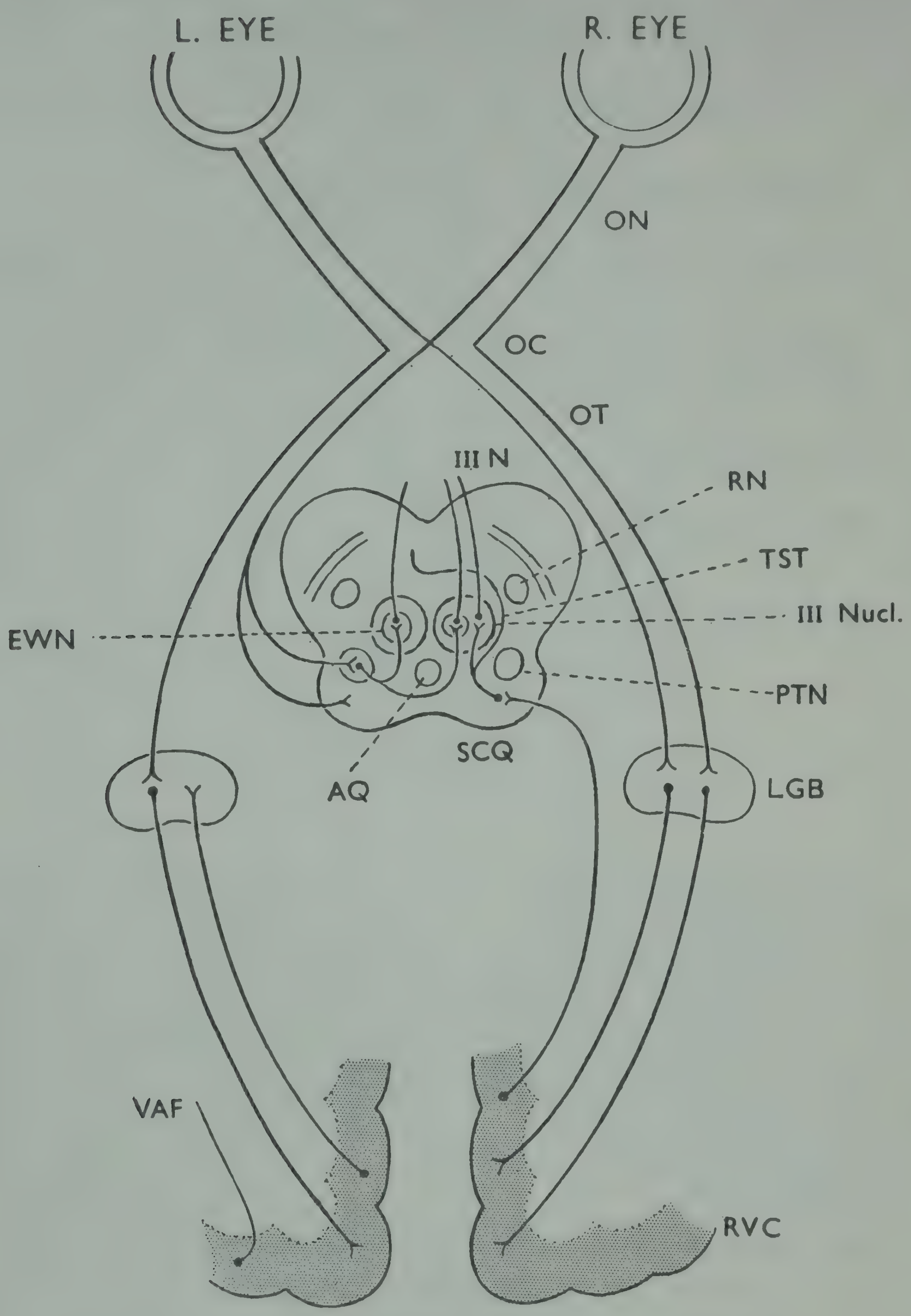

L. EYE
R. EYE
ON
OC
OT
III N
RN
TST
III Nucl.
EWN
PTN
SCQ
AQ
LGB
VAF
RVC

Fig. 49, 13

Diagram of the Visual Pathways

L. Eye = Left eye
R. Eye = Right eye
ON = Optic nerve
OC = Optic chiasma
OT = Optic tract
III N = IIIrd nerve
TST = Tectospinal tract
III Nucl = IIIrd nerve nucleus
AQ = Aqueduct of midbrain
RN = Red nucleus
EWN = Edinger-Westphal nucleus
PTN = Pretectal nucleus
SCQ = Superior corpus quadrigeminum
LGB = Lateral geniculate body
VAF = Visual association fibre
RVC = Right visual cortex

(By Courtesy of G. W. Pearce.)

nucleus on both sides from which nerve fibres pass via the ciliary ganglion to the sphincter pupillæ. The phenomenon of contraction of the pupil on accommodation but not to exposure to light, a common finding in certain diseases of the nervous system, is probably due to interruption of the pathways in the pretectal region. The light reflexes remain after cortical ablation since they do not require a cortical pathway. The very much longer sympathetic pathway to the eye described on p. 677 is illustrated in Fig. 49, 14.

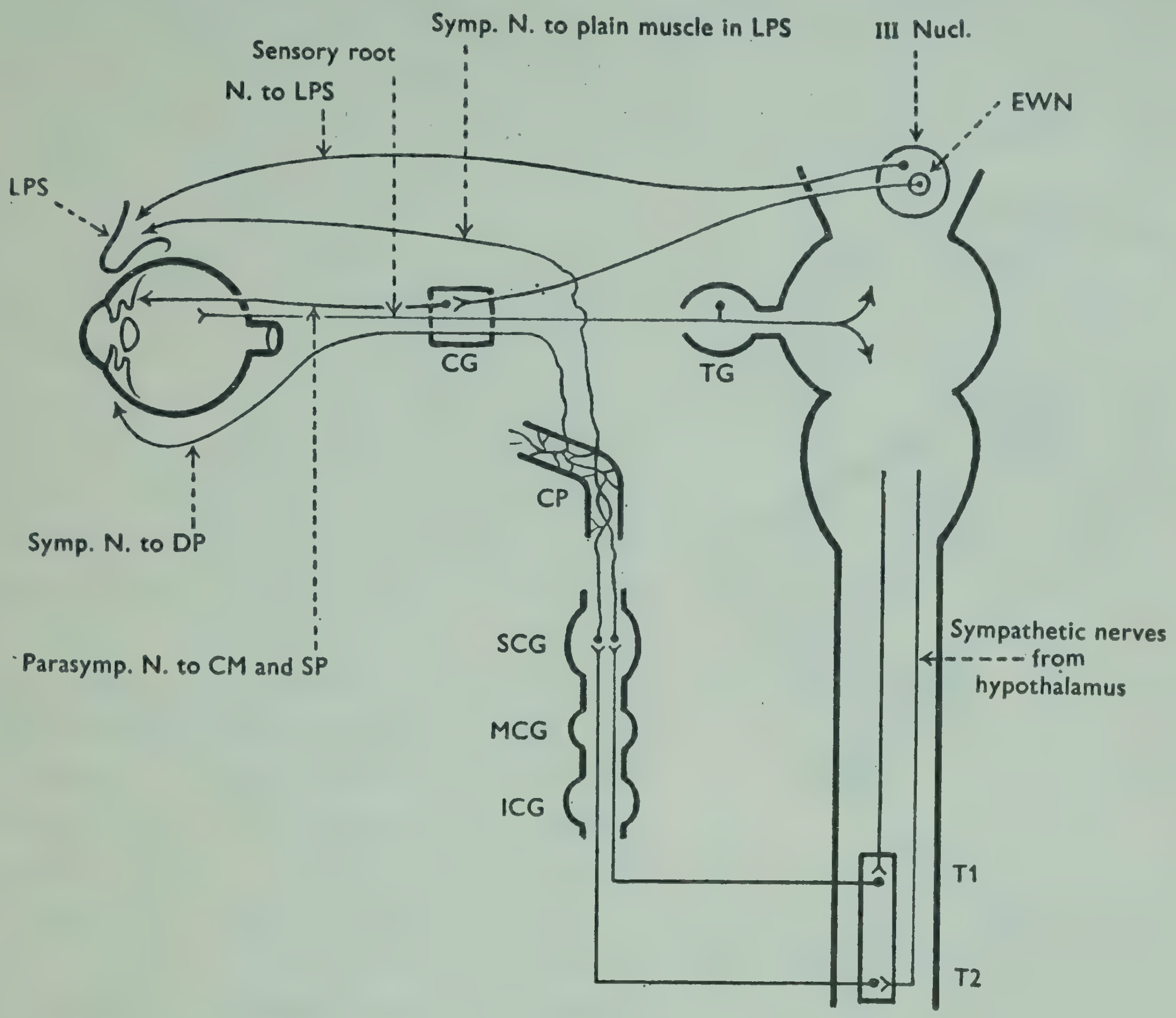

Symp. N. to plain muscle in LPS
III Nucl.
Sensory root
N. to LPS
EWN
LPS
CG
TG
CP
Symp. N. to DP
Parasymp. N. to CM and SP
SCG
MCG
ICG
Sympathetic nerves from hypothalamus
T1
T2

FIG. 49, 14

Diagram of the course of the autonomic fibres to the eye.

III Nucl. = oculomotor nucleus.
EWN = Edinger-Westphal nucleus.
LPS = Levator palpebrae superioris.
CG = Ciliary ganglion.
T1, T2 = First and second thoracic segments of spinal cord.
SCG, MCG and ICG = Superior, middle and inferior cervical sympathetic ganglia.
TG = Trigeminal ganglion.
DP = Dilatator pupillæ.
SP = Sphincter pupillæ.
CM = Ciliaris muscle.
CP = plexus on internal carotid artery.

(*By Courtesy of G. W. Pearce.*)

THE TEMPORAL LOBES

The route taken by impulses from the cochlea to the medial geniculate body and to a relatively small area of auditory cortex in the superior temporal convolution has already been described (p. 721). In the monkey, pure tones of high frequency applied to the ear produce nerve discharges in the caudal part of this area while low tones cause electrical disturbances in the rostral part of the same area (Fig. 49, 15). The cortical electrical response to any one tone may be reduced or abolished by sounding another tone; masking (p. 728) may thus be in part a cortical phenomenon.

In the cat destruction of one cochlea alone, or of one temporal lobe alone, causes a hearing loss of 3 to 5 decibels. If in addition to removing one temporal lobe the opposite cochlea is destroyed, the hearing loss when the crossed fibres only are intact increases to about 15 decibels, but the animal is by no means deaf. If after removal of one temporal lobe the cochlea on the same side is destroyed the loss in the presence of the uncrossed fibres only is again about 15 decibels. The crossed and uncrossed fibres from the cochlea have, therefore, in the cat about equal acoustic significance. After ablation of both temporal lobes in the monkey the behaviour of the animal is essentially normal and words of command are apparently understood; although localization of the source of a sound is poor, the animal is not deaf. Faradic stimulation of the superior temporal convolution in man (area 22) gives subjective buzzing, clicking, or booming sounds. Although ablation of one temporal cortex in man does not produce deafness, location of the source of sounds may be faulty. Furthermore it may, especially if it is on the left side, impair the understanding of the meaning of sounds and of spoken words (semantic or verbal aphasia).

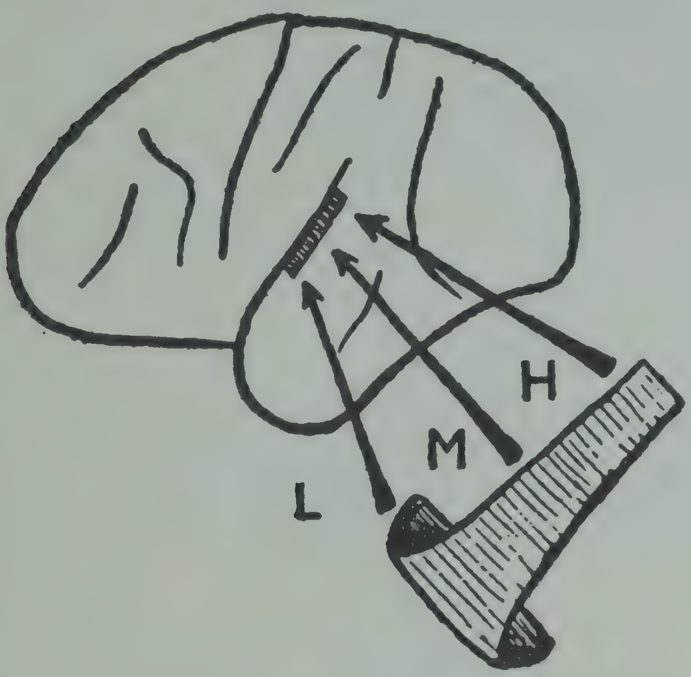


Fig. 49, 15

Representation of the basilar membrane of the cochlea in the auditory area on the temporal lobe of the monkey's brain. L, M, H indicate the positions for reception of nerve impulses due to tones of low, medium and high frequency (E. D. Adrian (1944), *Brit. med. J.* ii, 138).

Vestibular impulses also end in the temporal lobes but the pathway is not yet known. Dizziness occurs when certain parts of this cortical area is stimulated in man but removal of the temporal cortex does not produce any disturbances of equilibrium. It has been found impossible to assign any particular function to the lower parts of the temporal lobe.

The effects of lesions in the temporal lobes in man are extremely variable but in one variety the patient gets into a dreamy state and may have vivid memories of his past. Penfield, by stimulating points over a wide region of the temporal cortex in conscious patients, has evoked recollections which were sometimes fragmentary but were often elaborate and vivid, such as the recollection of a piece of music.

The Parietal Lobes

When the region of the postcentral gyrus in man (areas 3, 1 and 2) is stimulated electrically, sensations of touch and pressure are aroused. Occasionally feelings of warmth are elicited but never painful sensations, and no muscular movements occur if stimulation of the precentral gyrus is avoided. Sensation from the leg, trunk, arm and face on the opposite side of the body is represented in that order from above downwards and area representation in the sensory cortex (*somæsthetic area*) is, therefore, very similar to that of the motor cortex (area 4) lying immediately in front of it (Fig. 49, 16). The postcentral convolution receives impulses via the thalamus from the medial lemnisci and spinothalamic tracts. Stimulation of the posterior lip of the precentral gyrus, which is usually regarded as part of the motor area, also gives rise to sensations similar to those arising on stimulation of the postcentral gyrus.

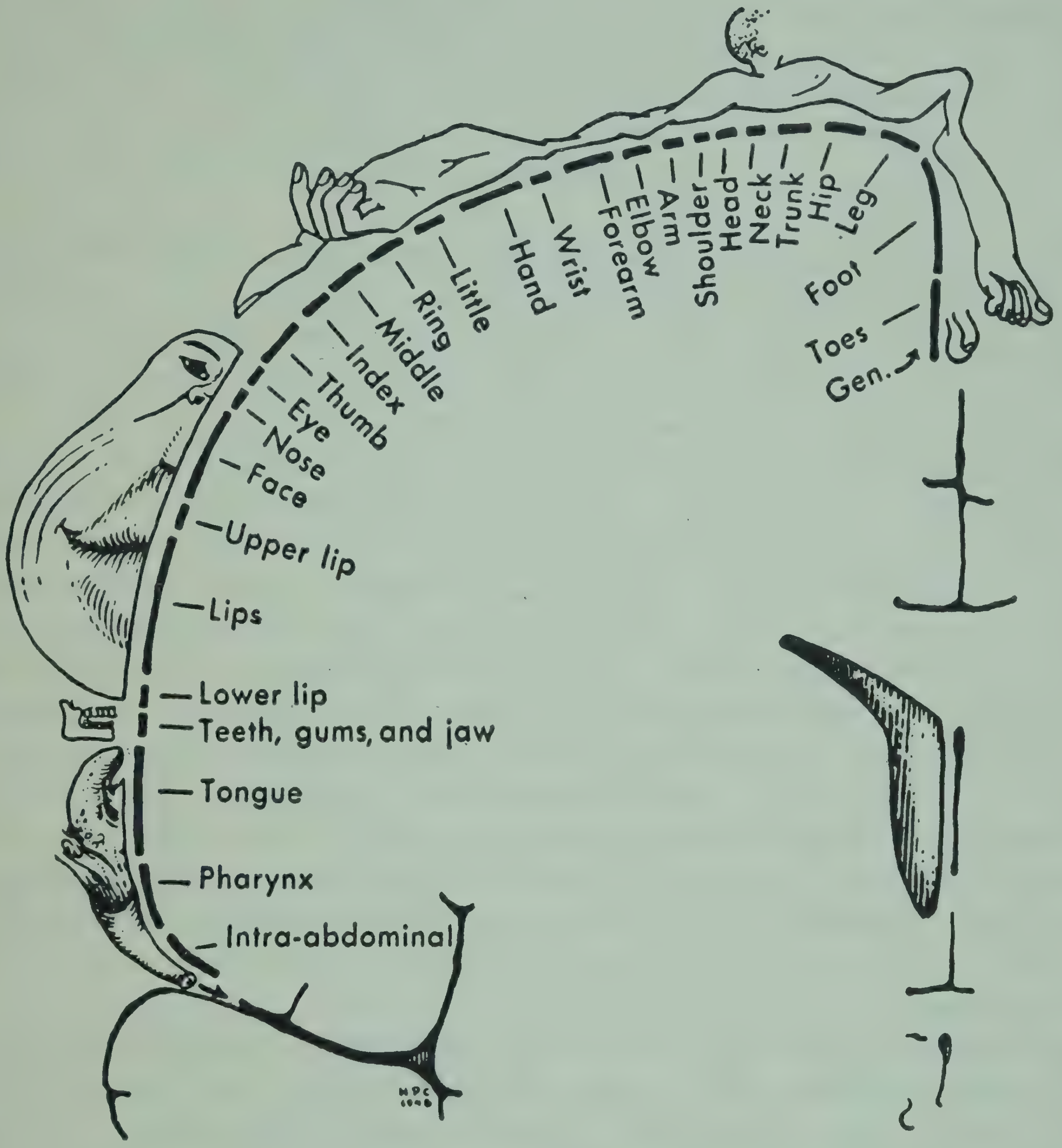

Little
Ring
Middle
Index
Thumb
Eye
Nose
Face
Upper lip
Lips
Lower lip
Teeth, gums, and jaw
Tongue
Pharynx
Intra-abdominal
Hand
Wrist
Forearm
Elbow
Arm
Shoulder
Head
Neck
Trunk
Hip
Leg
Foot
Toes
Gen.

FIG. 49, 16

Sensory homunculus. The right side of the figurine is laid upon a cross-section of the hemisphere, drawn somewhat in proportion to the extent of sensory cortex devoted to it. The length of the underlying block lines indicates more accurately the comparative extent of each representation (W. Penfield & T. Rasmussen (1950), *The Cerebral Cortex of Man*, p. 44. New York: Macmillan).

Recently an electrical method of mapping sensory areas in animals has been used by Adrian and others. Leads are taken from the cortex to an amplifier and thence to an oscillograph or a loudspeaker. The experimenter, by touching various parts of the animal's body in turn, finds the area of skin giving maximum cortical activity as indicated by the height of the oscillograph deflexions or by the noise in the speaker. In this way the representation of the body surface—leg, arm and one side of the face from above downwards—on the contralateral postcentral gyrus has been confirmed.

Removal of the postcentral gyrus leads to impairment, but not complete abolition, of the ability to detect differences in weight and texture and of stereognosis—the appreciation of form in three dimensions by the feel of an object. The latter implies a great deal more than the sense of touch since temperature and deep pressure sensibilities as well as appreciation of weight and position are all involved. It is concluded that this area of the cortex is capable of fine discriminations; that is to say, not only is the stimulus recognized but small differences in its intensity are appreciated as well. It is important to note that anæsthesia never results from a purely cortical lesion.

A lesion of the left angular gyrus, the middle part of the inferior parietal lobule, in a right-handed person causes sensory aphasia, that is to say although the patient can speak he talks nonsense; further, while he can perform the act of writing he does not write spontaneously. According to Head, large lesions in the posterior parietal lobe lead to a defect in word formation and difficulty in naming objects (nominal aphasia). Lesions in the parietal area in man may produce a loss of the ability to find the way even in the patient's own house together with a loss of the ability to conceive space even in two dimensions. When we are awake we are aware of the position of our body relative to other objects around us and of the position of its various parts, head, trunk and limbs, to one another; we are also aware of the changes in all these relationships. To this sort of awareness the term 'body-image' has been assigned. This image is, of course, dependent on a continuous stream of impulses from the proprioceptors and exteroceptors arriving at the brain. There is a certain amount of evidence that these impulses are co-ordinated in the parietal lobe, probably bilaterally. If this region is damaged then the patient may have difficulty in carrying out such everyday acts as dressing or sitting down in a chair.

Taste has recently been localized in the lowest part of the postcentral gyrus but the pathways from the tongue have not yet been fully worked out. Lesions in this area cause disturbances of taste as well as cutaneous sensory impairment on the face of the opposite side. Electrical stimulation of this area gives rise to taste sensations.

After removal or destruction of the postcentral gyrus there is a considerable recovery of sensation, a fact which may be explained by ipsilateral representation of sensation or by supposing that representation is not confined to the postcentral gyrus as is usually assumed. Sensation from deep structures is represented contralaterally but from cutaneous surfaces bilaterally. The thalami receive impulses from both sides of the body but each thalamus relays only to the ipsilateral cortex.

THE ELECTROENCEPHALOGRAM

Electrical potentials developed in the human brain can be detected on the surface of the head, and records of these potentials, obtained by means of a string galvanometer, were first published by Berger in 1929. Nowadays electrodes are placed at various positions on the scalp and connected through a valve amplifier to an ink-writing oscillograph. Alterations in electrical potential occur in a series of waves that are not quite regular although the dominant or α-rhythm is about 10 a second with an amplitude of about 50 microvolts. These regular changes of potential are produced mainly by the occipital cortex around the visual area (visual association area) and to a smaller extent from the frontal association areas. The α-rhythm is best seen when the subject is relaxed with eyes closed; when the eyes are opened and attention is paid to objects in the visual field, or when a mental task like mental arithmetic is undertaken, the rhythm disappears and is replaced by small rapid waves (Fig. 49, 17). The α-rhythm is therefore, as Adrian puts it, a "rhythm of inattention." The origin of the α-rhythm is a little difficult to decide upon; it can scarcely be the product of nerve impulses which are passing in all directions through the complex cortical interconnecting pathways in the brain in a manner unlikely to give rise to a series of fairly regularly occurring waves of electrical potential. Since such waves have been obtained in animals after section of the mid-brain, that is after removal of all afferents except the olfactory and optic nerves, they may represent the spontaneous activity of the cortical neurones but this matter is disputed. Delisle Burns has succeeded in isolating slabs of cerebral cortex in the cat by undercutting them while still preserving their blood supply. Although the nerve cells in the slab retained their excitability in the unanæsthetized animal they showed no spontaneous electrical activity. His experiments suggest that rhythmical cortical discharges can be maintained only if the cells are driven from a source outside the cortex, that is only if the cortical cells receive afferent impulses from other parts of the brain.

More rapid electrical oscillations are obtained further forward in the sensorimotor area. These β-waves have a frequency of 16 to 25 per second. δ-waves with a rhythm of about one per second are obtained from sleeping subjects (Fig. 49, 20).

Young children below eighteen months show electroencephalograms of low amplitude with little rhythmic activity. The adult electrical record is not usually reached till the age of seven or eight years when the α-rhythm becomes the dominant rhythm.

When the eyes are exposed to a uniform field lit with a flickering light the encephalographic potentials occur at the same rate as the flicker up to 30 per second (Fig. 49, 18), as if the cells in the visual cortex were discharging synchronously in response to each flash of light. This rhythm spreads to the neighbouring association regions. This 'photic stimulation' can cause changes in the mental state and may even induce an epileptic attack.

In hypoglycæmia the waves tend to become slower and larger (Fig. 49, 19). In epilepsy, massive discharges of cortical neurones produce much larger waves of alteration of potential which also last much longer than normal. However, similar electroencephalograms have been recorded in about 10 per cent. of apparently healthy persons.

The electroencephalogram has proved of value in the diagnosis of epilepsy and intracranial disease and in defining the area affected by a pathological process. The contraction e.e.g. for electroencephalogram is in current use.

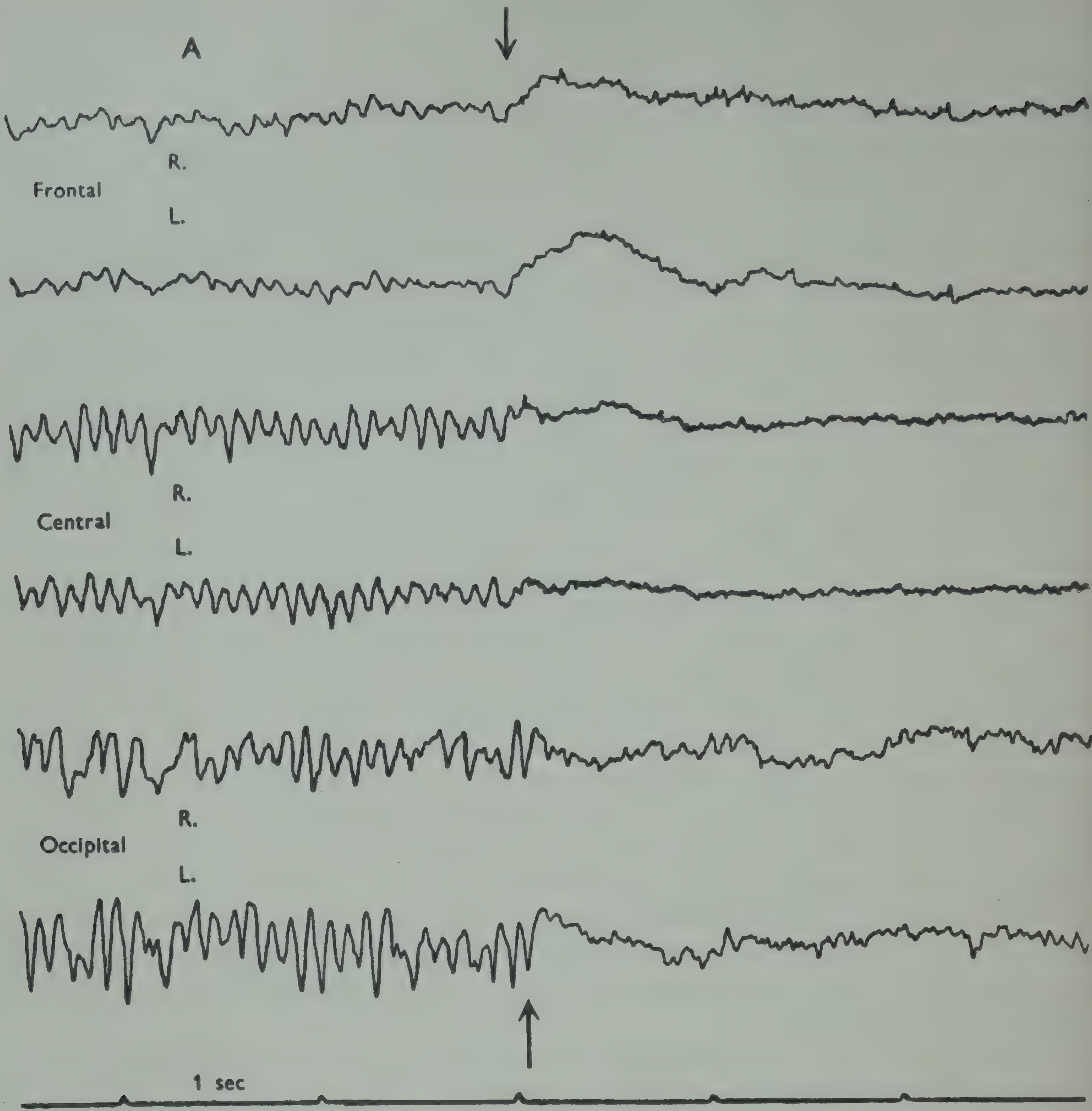


FIG. 49, 17

The normal human electroencephalogram (e.e.g.) as seen in bipolar antero-posterior electrode placements. The dominant rhythm (11 c/s) spreads forward from the occipital regions. This α-rhythm is inhibited by eye opening which occurred at the arrow. All six traces were recorded simultaneously.

(*By courtesy of D. A. Pond.*)

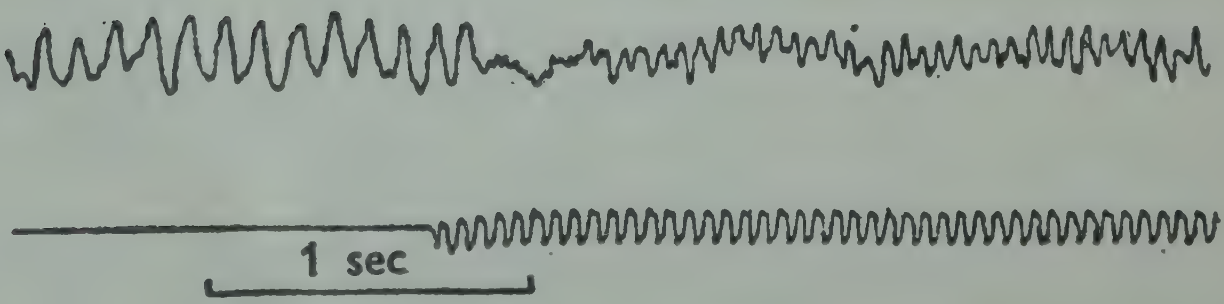
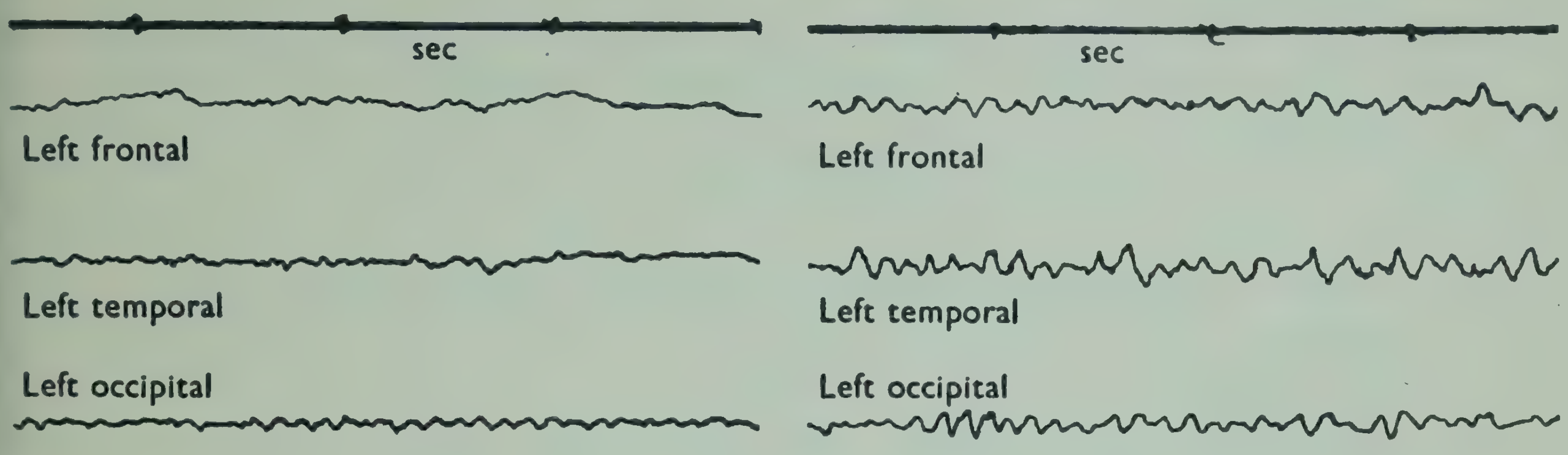


FIG. 49, 18

Electroencephalograms from the occipital region in man, showing the effect of flickering light. Above, e.e.g.; below, record from photoelectric cell showing rate of flicker (18 per sec); the eyes were opened as the flicker began (E. D. Adrian (1943), *Lancet*, ii, 33).

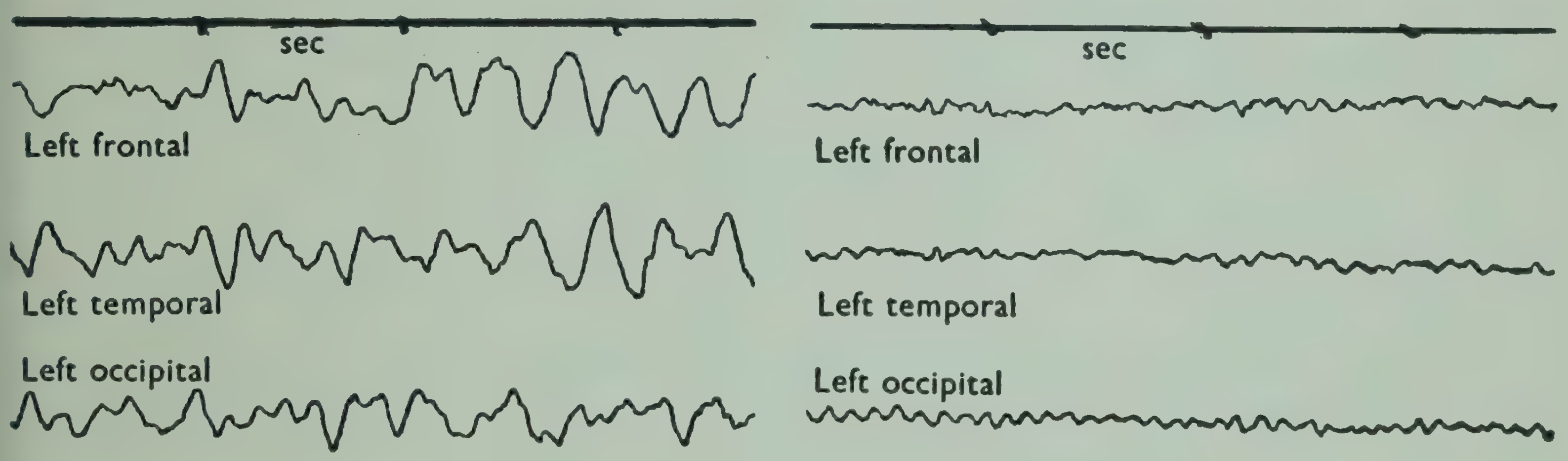
SLEEP

The nervous system cannot long continue to function without sleep, which is distinguished from other forms of unconsciousness by the relative ease with which the waking conscious state is reached. Sleep is no doubt a cerebral phenomenon but the old notion that it is brought about by cerebral anæmia can no longer be maintained; the total blood flow through the brain of man during sleep is not altered although there may, of course, be local alterations of blood flow that cannot be measured. The arterial oxygen saturation has been measured during sleep by an oximeter (p. 573) clipped to the subject's ear; in deep sleep the arterial saturation may fall from 96 per cent. to 87 per cent. Whether this is a cause or a result of sleep it is not possible to say. In sleep the pupils are small and the eyelids motionless; in feigned sleep the pupils are dilated and there are usually small movements of the eyelids.



1. Fasting blood sugar 109 mg. per 100 ml.

2. 13 min after injection of insulin. Blood sugar 55 mg. per 100 ml.



3. 20 min after injection. Blood sugar 40 mg. per 100 ml.

4. After intravenous glucose

FIG. 49, 19

Sequence of changes in human electroencephalogram during hypoglycæmia following intravenous insulin (0·5 unit per kg. of body weight). Note that the appearance of slow activity in record 2 occurs first in the frontal regions and spreads to involve the whole cortex as the blood sugar falls (record 3). Complete and immediate recovery occurs following intravenous glucose. Time in minutes following insulin injection. All records are from the left frontal, temporal and occipital regions, the electrodes being in pairs along the head (bipolar recording).

(*By courtesy of D. A. Pond.*)

Pavlov tried to explain sleep as the effect of internal inhibition (p. 914), but it cannot be entirely a cortical phenomenon since decorticate animals, as well as very young animals and human infants, show rhythmic waking and sleeping which, however, bear no relationship to day and night but are repeated several times during the twenty-four hours. The ability to sleep during the night and remain awake during the day is probably dependent on the development of conditioned reflexes.

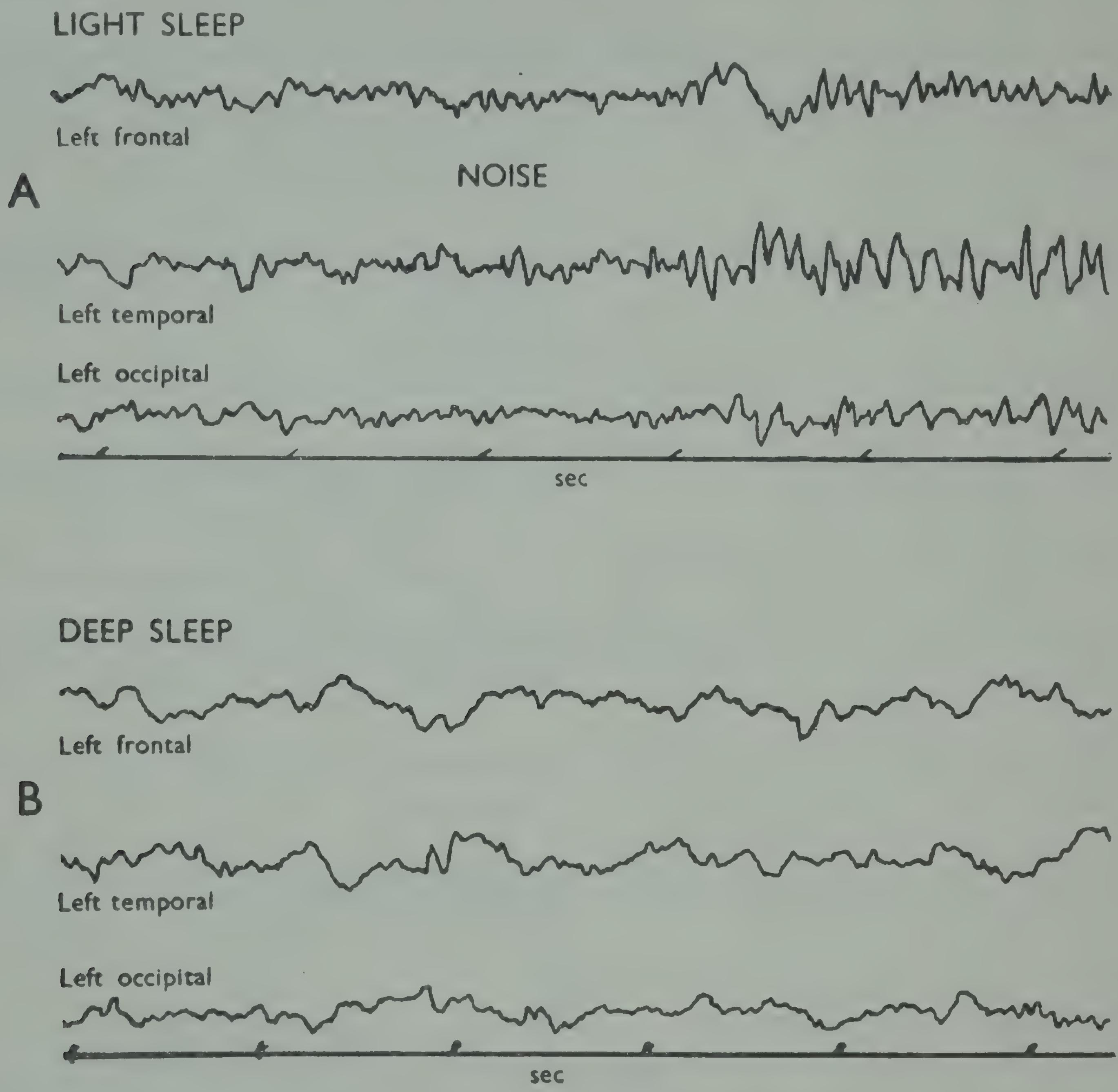


Fig. 49, 20

The electroencephalogram in sleep. The first channel records from the left frontal, the second from the left temporal and the third from the left occipital region. In record A a noise caused a return to the waking state. Record B in deep sleep shows the dominant activity to have shifted from the α-band to the slower δ bands; the very slow irregular waves have a frequency of 1 c/s or less.

(*By courtesy of D. A. Pond.*)

According to Bremer, a cat with its mid-brain sectioned immediately behind the third nerve nucleus, the basilar artery remaining intact, exhibits sleep. Just as in deep sleep, the pupils become very narrow and the eyes look down and the electroencephalogram shows a rhythmic waxing and waning of the α rhythm. A cut made between the medulla oblongata and the spinal cord leaves the brain 'awake'. These findings led Bremer to the conclusion that afferent impulses reaching the fore-brain maintain wakefulness. Magoun believes that the cortex is kept awake, not by impulses travelling in the classical sensory paths of the lemnisci to the thalamus and cortex, but by impulses travelling up the central brain stem reticular formation to the posterior hypothalamus and the medial nuclei of the thalamus and then to the cortex. Magoun showed that stimulation of the reticular formation and the diencephalic region just mentioned gave the e.e.g. picture of arousal with fast low amplitude waves but that lesions in the same area produced a state of sleep associated with a characteristic e.e.g. sleep record of high amplitude slow waves. Magoun showed further that lesions involving all the lemniscal tracts but sparing the central reticular formation did not produce sleep. Moreover, sleep was produced by injury to the reticular formation which spared the lemnisci.

Prolonged sleep (hypersomnia) is associated with lesions of the posterior hypothalamus such as are encountered in encephalitis lethargica (sleepy sickness), and injection of fluid into this area, if it produces mechanical injury, also causes sleep. Hess found that stimulation of the posterior hypothalamus caused arousal, and from the same area he also obtained increased motor activity and excitability, dilatation of the pupils and a rise in blood pressure and heart rate. Such observations support the conception of a 'waking centre' in the posterior hypothalamus. Hess has produced 'sleep' by stimulation of the thalamus in the region of the interthalamic connexion.

It will be appreciated that in all such experiments it is not possible to say in a direct way that an animal is truly asleep, and the experimenter is always unable to say whether an animal is exhibiting true sleep or whether it is merely showing a great decrease of activity.

Sleep occurs most readily when there is fatigue after muscular exertion; it has been suggested on rather doubtful evidence that a soporific substance produced during wakefulness is responsible for the onset of sleep.

For experimental purposes the depth of sleep is usually gauged by finding the minimum noise necessary to wake the subject. In healthy adults there are at least two main types of sleep behaviour. The commonest pattern is two periods of deep sleep each lasting several hours separated by a period of relatively light sleep occurring in the early hours of the morning, during which the subject moves about more in bed, may become restless and disturbed and even momentarily awaken. Many persons, however, have only one of these periods of deep sleep, the commonest variation being the one in which sleep is deepest from one to two hours after it begins, becoming gradually lighter until the subject awakens at his normal time. The best method of inducing sleep is to cut off afferent stimulation as much as possible by darkening the room and excluding noise. Warmth and a comfortable bed, by promoting muscular relaxation, reduce proprioceptive afferent impulses and facilitate sleep.

REFERENCES

Adrian, E. D. (1943). Sensory areas of the brain. *Lancet*, ii, 33–35.

Adrian, E. D. (1944). Brain rhythms. *Nature*, **153**, 360–362.

Brain, W. R. (1950). The cerebral basis of consciousness. *Brain*, **73**, 465–479.

Brain, W. R. (1952). The contribution of medicine to our idea of the mind. *Rede Lect.* Cambridge University Press.

Bremer, F. (1953). *Some Problems in Neurophysiology.* University of London, Athlone Press.

Bucy, P. C. (Ed.) (1944). *The Precentral Motor Cortex.* Urbana, Illinois: University of Illinois Press.

Cairns, H. (1952). Disturbances of consciousness with lesions of the brain stem and diencephalon. *Brain,* **175,** 109–146.

Carmichael, E. A. (1953–1954). Hemispherectomy and the localization of function. *Lect. Sci. Basis Med.* **3,** 93–104.

Critchley, M. (1953). *The Parietal Lobes.* London: Arnold.

Druckman, R. (1952). A critique of 'suppression' with additional observations on the cat. *Brain,* **75,** 226.

Eccles, J. C. (1953). *The Neurophysiological Basis of Mind.* Oxford: Clarendon Press.

Fulton, J. F., Aring, C. D. & Wortis, S. B. (Edrs.) (1948). The frontal lobes. *Res. Publ. Ass. Nerv. Ment. Dis.* **27.** Baltimore: Williams & Wilkins Co.

Hill, D. & Parr, G. (Edrs.) (1950). *Electroencephalography.* London: Macdonald.

Lassek, A. M. (1954). *The Pyramidal Tract. Its Status in Medicine.* Springfield: Thomas.

Liddell, E. G. T. (1953). The brain and muscle management. *Quart. J. exp. Physiol.* **38,** 125–137.

Liddell, E. G. T. & Phillips, C. G. (1950). Thresholds of cortical representation. *Brain,* **73,** 125–140.

Magoun, H. W. (1950). Caudal and cephalic influences on brain stem reticular formation. *Physiol. Rev.* **30,** 459–474.

Magoun, H. W. & Rhines, Ruth (1947). *Spasticity. The Stretch Reflex and Extra-pyramidal Systems.* Springfield, Illinois: Chas. C. Thomas.

Peele, T. L. (1954). *The Neuroanatomical Basis for Clinical Neurology.* London: McGraw-Hill.

Penfield, W. & Faulk, M. E. (1955). The insula: further observations on its function. *Brain* **78,** 445–470.

Penfield, W. & Rasmussen, T. (1950). *The Cerebral Cortex of Man.* London: Macmillan.

Penfield, W. & Welch, K. (1951). The supplementary motor area of the cerebral cortex. *Arch. Neurol. Psychiat.* **66,** 289–317.

Richter, D. (Ed.) (1950). *Perspectives in Neuropsychiatry.* London: Lewis.

Symposium on functional organization of the cerebral cortex (1947). *Fed. Proc.* **6,** 437–460.

Walshe, F. M. R. (1948). *Critical Studies in Neurology.* Edinburgh: Livingstone.

Walshe, F. M. R. (1951). On the interpretation of experimental studies of cortical motor function: with special reference to the 'operational view' of experimental procedures. *Brain,* **74,** 249–256.

CHAPTER 50

CONDITIONED REFLEXES

THE methods for exploration of the function of the cerebral cortex which have been given in the previous chapter nearly all involve more or less severe surgical procedures. The method discussed in this chapter provides a fundamentally different approach to the problem because the behaviour of the conscious intact animal is the subject of study.

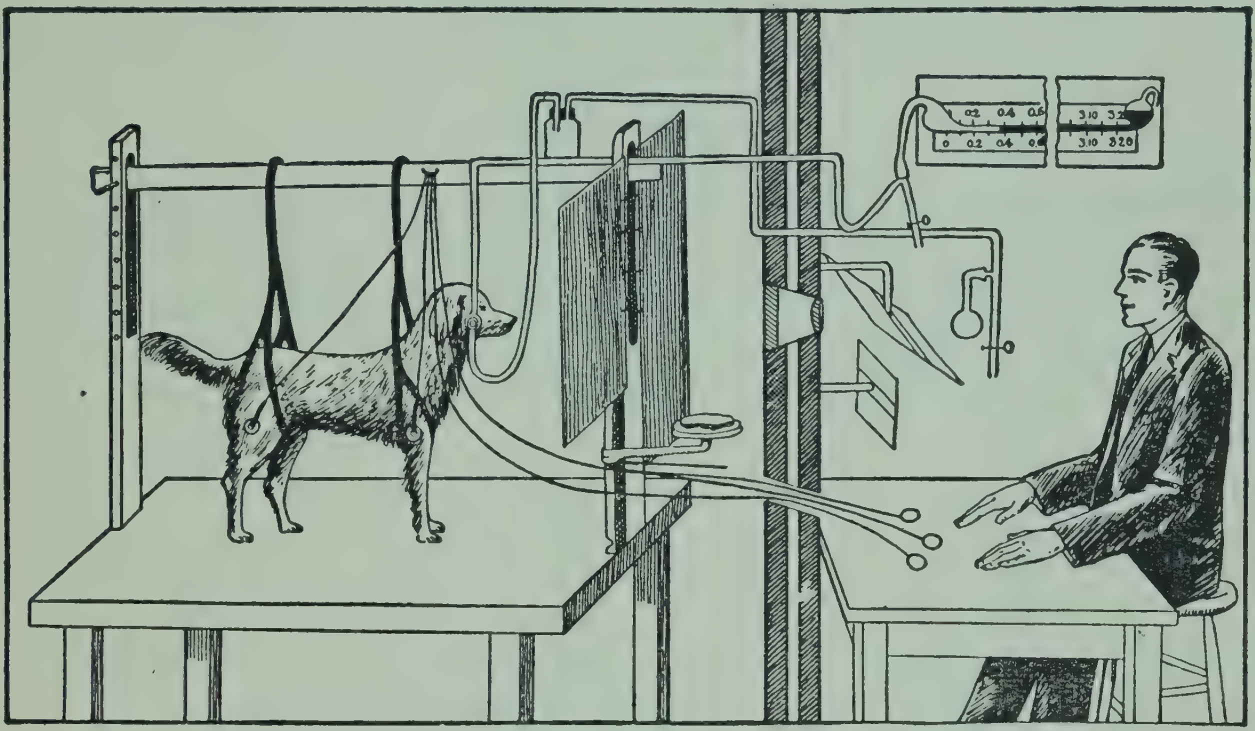


FIG. 50, 1

A diagram of the double chamber used in the investigation of conditioned reflexes. The animal stands on a table in a loose harness. A hemispherical bulb placed over the parotid fistula has two outlet tubes passing into the observer's room. One of the tubes is connected to a graduated tube for measuring the saliva and the other goes to a hand pump used to apply negative pressure to draw off the saliva. The operator with his various controls occupies the right-hand chamber from which he can observe the animal through a periscope (Pavlov, I. P. (1928), *Lectures on Conditioned Reflexes*, p. 271. London: Lawrence & Wishart).

At the end of the nineteenth century the Russian physiologist Pavlov was engaged in studying problems of digestion and in particular the problem of the control of salivation. It was soon found that this was not a simple physiological matter, for a large variety of factors, such as for example the sight, as well as the presentation, of food would cause an animal to salivate and the experimental results were considerably influenced by the presence and behaviour of the observer. Pavlov, realizing the importance of these extraneous factors, gave up his work on the digestive glands and began a long investigation of "conditioned reflexes" which, as he said, opened the way to a study of the "whole indivisible life."

Pavlov used dogs in his experiments and, to avoid disturbing influences, carried out the work in quiet rooms specially built for this purpose. The animals were trained to stand still on a table, their movements restricted only by loops passing loosely round the legs to discourage struggling (Fig. 50, 1). The experimenter occupied another room from which he could observe the animal's behaviour without being seen, and from which by remotely controlled apparatus he could present food to the animal or apply various stimuli. A tube from the dog's exteriorized parotid duct was carried through into the observer's room so that the production of saliva could be measured.

As the result of experiments made in this way Pavlov classified all reflex behaviour into two categories: (1) *unconditioned* or inborn and (2) *conditioned* or acquired. If food is placed in an animal's mouth salivation occurs. This reflex, present in all animals, was the unconditioned reflex most studied by Pavlov. The presence of food in the mouth is the unconditioned stimulus. A neutral stimulus, such as the ringing of a bell, which normally evokes no salivary response, can be presented at the same time that the animal is given food. When this combined stimulation has been repeated a number of times it is found that the neutral stimulus given alone now evokes a flow of saliva. The ringing of the bell, originally a neutral stimulus, has thus acquired some of the properties of the unconditioned stimulus; it is now referred to as the 'conditioned' stimulus. The animal has acquired a new reflex which is described as a 'conditioned' reflex.

TABLE 50, 2

Development of a Conditioned Reflex to a Sound of 637·5 *c/s.*

Number of times combination of sound with feeding had been performed	Extent of reflex in drops of saliva per 30 seconds	Latent period of conditioned reflex in seconds
1	0	..
9	18	15
15	30	4
31	65	2
41	69	1
51	64	2

(From G. V. Anrep (1920), *J. Physiol.* **53,** 367.)

The unconditioned and conditioned stimuli must be presented *together* for a certain number of times before the conditioned reflex is established. This is the process of training or *reinforcement*. Thus, in the particular experiment summarized in Table 50, 2, after nine repetitions of the combined stimuli the conditioned stimulus (the sound) was presented alone and the amount of salivation recorded; the strength of the conditioned reflex was tested in the same way after 15, 31, 41 and 51 repetitions of the combined stimuli. It will be seen that the reflex had become established both in regard to extent and to latent period when sound and food had been presented together 31 times. This is an example of a *simultaneous conditioned reflex*. Although the reflex persists for a long time the extent of the response to the conditioned stimulus tends to decline. It can, however, be reinforced in the same way that it was established, that is by the simultaneous presentation of the conditioned and the unconditioned stimuli. On the other hand, if the conditioned stimulus is repeatedly presented without the unconditioned stimulus (that is, sound but no food) then the conditioned reflex is *extinguished*.

A recently established conditioned reflex is unstable and may be inhibited by external stimuli which have the effect of distracting the animal's attention or which lead, as Pavlov put it, to an investigatory reflex. This is illustrated in the experiment described in Table 50, 3. A conditioned reflex was first set up to a visual stimulus and, after training, a relatively constant response, as judged by the flow of saliva, was obtained when the visual stimulus was presented. The strength of the reflex was then tested by presenting the conditioned stimulus (the light) accompanied by the playing of a gramophone record. The first playing of the record led to almost complete inhibition of the reflex (amount of saliva reduced to one-tenth); later the effect of the gramophone became less. This kind of inhibition of a conditioned reflex is called *external inhibition.*

TABLE 50, 3

External Inhibition

	Per cent.
Initial strength of conditioned reflex	100
Strength of reflex at first application of gramophone record . .	10
,, ,, second ,, ,, ,, . .	50
,, ,, third ,, ,, ,, . .	65
,, ,, fourth ,, ,, ,, . .	85
,, ,, fifth ,, ,, ,, . .	90
,, ,, sixth ,, ,, ,, . .	94
,, ,, seventh ,, ,, ,, . .	100

(From C. Lovatt Evans (1928), *Recent Advances in Physiology*, p. 249. London: Churchill.)

If a conditioned reflex is repeatedly extinguished by very frequent presentation of the conditioned stimulus without the unconditioned (for example by ringing the bell without giving food) it does not simply disappear but it becomes a *negative conditioned reflex.* This phenomenon is illustrated by the following experiment. When a reflex set up to the sound of a bell is extinguished by frequent ringing of the bell without giving food then eventually no salivation will be obtained when the bell is rung. If at this stage the bell is rung and a gramophone record is played, salivation (a positive response) will occur. The distraction (external inhibition) provided by the gramophone has inhibited the negative conditioned reflex and a positive response has been obtained.

In the early stages of establishing a conditioned reflex to a sound, for example the click of a metronome, any other sound may elicit salivation, but after a time no sound differing greatly from that of the metronome can cause salivation. Similarly, if a conditioned reflex is set up to a note of a particular frequency then it will be found that a note of a neighbouring frequency can cause salivation, although not necessarily to the same extent. This property of a conditioned reflex, referred to as *specificity*, can be greatly increased in the following way. Tone A is always accompanied by food whereas tone B of a different frequency never is; tone A when given alone always produces salivation. Tone B alone may at first produce salivation but when training has continued for some time tone B alone ceases to cause salivation. In other words the animal is able to distinguish between the two tones A and B. In this case tone B has become something more than a neutral stimulus since it produces a state of *internal inhibition.* This can be demonstrated by the inhibitory effect of tone B on other conditioned reflexes in the same animal. The limits of discrimination may be discovered by bringing the frequency of tone B nearer and nearer to that of tone A. By the use of this method it has been found that the dog can distinguish very small intervals of pitch (an eighth of a tone) and that it can hear frequencies up to 60,000 c/s. By similar experiments with objects placed on the skin it has been shown that differences of temperature of 1° C can be recognized. The dog can readily distinguish between different shapes and varying degrees of luminosity, although he is unable to distinguish colours if very great care has been taken to make their luminosity equal.

The technique of establishing a conditioned reflex can be varied in several ways. For example, a whistle may be blown continuously for sixty seconds immediately before the animal is fed. When the whistle is blown without feeding (to test the conditioned reflex) it is found that salivation occurs one minute after the beginning of the whistling. This is an example of a *delayed conditioned reflex.* If the attention is distracted, for example by playing a gramophone record, at any time up to sixty seconds after blowing the whistle, a flow of saliva is apt to occur before the sixty second delay period has expired. This was interpreted by Pavlov to mean that the distraction inhibited the state of internal inhibition which existed during the delay interval. If the unconditioned stimulus, that is the food, is given first and the neutral stimulus follows, a conditioned reflex cannot be established. Events occurring after the food has been received are of no value as conditioning stimuli.

It is also possible to establish by the following type of experiment what has been called a *trace conditioned reflex.* During training a note is sounded for a short time, and after a pause of fixed duration the animal is fed and salivates at once. When this procedure has been repeated a number of times the note, the conditioned stimulus, is sounded alone. After an interval, equal to the pause used during training, salivation occurs. It is clear that salivation is produced not as a direct response to the conditioned stimulus but indirectly through some representation or *trace* of it in the central nervous system. There is, as might be expected by analogy with the delayed conditioned reflex, a state of internal inhibition in the interval between the conditioned and unconditioned stimuli. When establishing trace conditioned reflexes Pavlov found that if the interval between the conditioned and unconditioned stimulus was very long, say half an hour, the animal tended to go to sleep in the interval between the two stimuli and then it wakened up and salivated. Although it is quite reasonable to suppose that the rhythm of sleep—the feeling of tiredness at a given time and the wakening at another nearly fixed time—is a conditioned phenomenon, the notion that sleep is the result of prolonged internal inhibition of a large area of the brain, as suggested by Pavlov, is open to many objections.

Conditioned reflexes have been established to all sorts of stimuli and with many different unconditioned reflexes as a basis, for example salivation, pupillary reflexes, vasomotor reactions, vomiting, the knee jerk. Conditioned reflexes themselves have been used as a basis for building conditioned reflexes as described on p. 916.

Pavlov considered that in investigating conditioned reflexes he was studying the activities of the cerebral cortex. The name *analyser* was given to the peripheral end-organ together with its afferent nerves and the area of cortex receiving the afferent nerve impulses. For example, the auditory analyser consists of the organ of Corti, which as we already know makes an elementary analysis of sounds, the auditory nerve and the temporal lobes of the cerebral cortex. If, after a conditioned reflex has been established to an auditory stimulus, both temporal lobes are removed the reflex is not completely abolished but is greatly impaired and the power of discrimination is much reduced. It was supposed that the survival of such rather crude reflex activities was due to the fact that the analyser is not confined strictly to the temporal lobes. It is now known, however, that conditioned reflexes can be established after complete removal of the cerebral cortex (decortication). Decorticate dogs show quite abnormal behaviour. One such dog walked into obstacles as if unaware of their existence, ate food when it was put into his mouth but failed to find food even when he trod on it. His reactions to substances with strong tastes were normal; thus he rejected meat contaminated with quinine. It was found possible to establish a conditioned salivary reflex in this dog to the sound of a metronome, but only after 200 combinations of food and metronome; he also acquired a conditioned reflex to the sound of footsteps approaching his cage which indicated the arrival of food. Other experiments with decorticate dogs have been reported, but in none was there any refinement with practice and the movements were never so precise as in the intact animal.

The hypothalamus has recently been credited with control over many functions of the body but it has not been possible to set up conditioned reflexes involving this area. Conditioned reflexes involving the cerebellum have, however, been established.

Pavlov's experimental method has been the subject of much criticism. Admittedly the experimental conditions are abnormal. The dog stands motionless in the apparatus and reacts only to stimuli applied by the experimenter, and its behaviour is as a rule completely stereotyped. This, however, is not always so, as is indicated by the following description of an experiment by Liddell: "A dog had been conditioned to food by the sound of a metronome placed on a shelf above its head and by the hammer of a door-bell tapping on the under side of the table on which it stood. Occasionally during intervals between conditioned stimuli it peered over the edge of the table as if reminded of the tapping, or looked up toward the metronome, wagging its tail, licking its chops and salivating." It is difficult to fit the behaviour of this animal into the description of a conditioned reflex; it would ordinarily be described as spontaneous behaviour.

Bearing in mind these descriptions of conditioned reflexes in dogs we may be tempted to inquire to what extent similar phenomena can occur in man. Examples of human behaviour can be found which closely resemble the animal experiments in that it is easy to recognize conditioned and unconditioned stimuli. For example a patient who has an attack of asthma brought on by the sight or smell of flowers can develop a typical attack at the sight of artificial flowers.

Conditioned reflexes can be established in man under experimental conditions and Hudgins has built up reflexes which enabled his subjects to achieve voluntary control over their pupillary reflexes. When a strong light is shone into the eye the pupil becomes smaller—this is the basic unconditioned reflex. If a bell is sounded whenever the light is shone into the eye then the bell becomes a conditioned stimulus and the sound of the bell alone ultimately produces constriction of the pupil. The subject is now given a hand dynamometer (two bars kept apart by springs) with electrical switches so arranged that when he grasps it firmly and squeezes the bars together he closes the light and bell circuits; when he relaxes the circuits are broken. At first the subject squeezes at the command of the experimenter but later the bell and dynamometer are removed so that eventually the experimenter's commands to contract and relax are the conditioned stimuli which produce contraction and relaxation of the pupil. In the next stage of training the subject is asked to speak the words "contract" and "relax" and then to whisper them, and then finally to say them to himself. In this way the subject acquires voluntary control over his pupillary reflexes. Such an experiment, although fairly elaborate in its building up of one conditioned reflex upon another, is not to be regarded as freakish as it has been carried out on numerous subjects. This kind of experiment breaks down any rigid distinction between what is usually described as voluntary behaviour and that described as reflex or involuntary, and suggests that we ought to discard the idea that they are essentially different.

If we reflect on the experiments described in this chapter we find that the distinction between bodily and mental processes becomes indefinite. If in addition we make the justifiable assumption that all reflex activities can be conditioned, it is not difficult to imagine how mental processes can cause disorder of bodily function. Indeed many people are of the opinion that certain diseases, including even such well-defined pathological processes as peptic ulcer and exophthalmic goitre, may be the end-results of a prolonged disturbance of function produced primarily by psychic causes. This is the field of psychosomatic medicine.

EXPERIMENTAL NEUROSIS

The results of an experiment in which a dog was required to distinguish between an oval and a circle were described by Pavlov in 1927. A luminous circle was the signal for food and an oval of equal area and brightness indicated no food. When there was a considerable difference between the circle and the oval they were easily distinguished, but when the semi-axes of the oval were in the ratio 8 to 9, that is to say when the oval was very nearly a circle, the dog was unable to distinguish between them and there was a sudden change in its behaviour. It salivated profusely when either the circle or the oval was shown and it attempted to escape from its loops. It was not possible to use this animal for further conditioning until it had been rested for a long time.

In later work many instances of abnormal behaviour were seen, for example chronic somnolence or persistent immobility in which food was held in the mouth without being chewed or swallowed. During training paradoxical results were sometimes obtained, a weak conditioned stimulus for example evoking a greater salivation than a strong stimulus. Pavlov called these states experimental neuroses. The fixed position within the training frame and the monotonous repetitive routine of conditioning are partly responsible for the production of neuroses. They are most likely to develop where progressively finer discrimination is being carried out or where successively longer delayed responses are being established.

Neurosis can occur in dogs quite apart from experimental procedures. Pavlov described how some of his dogs in which conditioned reflexes had been patiently built up were trapped accidentally in their kennels in the Leningrad flood of 1924. When they were eventually rescued they were swimming about in their cages and the water was still rising. Some of them became stuporose after they were rescued and when this had passed off it was found that these dogs had lost their conditioned reflexes.

REFERENCES

Frolov, Y. P. (1937). *Pavlov and his School. The Theory of Conditioned Reflexes.* London: Kegan Paul.

Hudgins, C. V. (1933). Conditioning and the voluntary control of the pupillary light reflex. *J. gen. Psychol.* **8**, 3–51.

Konorski, J. (1948). *Conditioned Reflexes and Neuron Organization (a Criticism of Pavlov's Theories).* Cambridge University Press.

Pavlov, I. P. (1927). *Conditioned Reflexes*, translated by Anrep, G. V. Oxford University Press.

Russell, R. W. (1953–4). Experimental psychopathology. *Lect. Sci. Bas. Med.* **3**, 372–390.

CHAPTER 51

THE ENDOCRINE GLANDS

THE term *hormone* is applied to a chemical substance which, produced in one part of the body, enters the circulation and is carried to distant organs and tissues to modify their structure and function. The word, originally introduced in 1902 by Bayliss and Starling to describe the substance secretin (p. 254), is derived from a Greek word meaning to excite; however, we know now that not all such chemical messengers have a stimulating effect, for example enterogastrone (p. 245) and urogastrone inhibit gastric secretion and motility. As a rule the production of a particular hormone or group of hormones is confined to a single organ; for example thyroxine is produced only in the thyroid gland and this organ is then described as an endocrine gland or, since the hormone is not conveyed along a duct but passes directly into the blood, as a ductless gland or gland of internal secretion. The elaboration of specific hormones is not, however, confined to the endocrine glands, for both secretin and cholecystokinin for example are regarded as hormones. Further, noradrenaline and acetylcholine liberated at nerve-endings by the nerve impulse, though very rapidly destroyed at the sites of their production, may enter the circulation in sufficient amounts to produce effects in distant parts.

Most of the hormones have now been isolated in the pure state and some, including adrenaline and noradrenaline, thyroxine and certain hormones of the adrenal cortex, pituitary and gonads have been synthesized. It is still necessary in many instances to use biological rather than chemical methods for assaying the potency of extracts and preparations made from endocrine glands and such will be briefly mentioned in the sections that follow. It should be emphasized that the hormones, with few exceptions, are not species specific; extracts of sheeps' thyroid glands can increase the metabolic rate of guinea-pigs or stimulate the development and metamorphosis of tadpoles, and bovine insulin is used in the treatment of human diabetes mellitus. Many hormones are either proteins or linked to protein molecules and this means that they may be antigenic (p. 378). In other words, if the protein is 'foreign,' antibodies may be produced and these may have the effect of neutralizing or counteracting the action of the hormone; the serum behaves, then, as though it contained an anti-hormone.

The physiological actions of the various hormones have been explored in a number of ways. First by studying the effects of complete removal of an endocrine gland from the body. Secondly by grafting experiments or injections of active extracts of the gland to try to restore the animal to normal. Thirdly the effects of the administration of large amounts of the hormone to normal animals have been observed. Conclusions based upon the results of such experiments have in many cases been amplified by the study of disease in man. Disorders due to overactivity or to destruction of the various endocrine glands are well recognized, for it is now generally agreed that the only disturbances of secretion are excessive or diminished production and do not involve the elaboration of compounds of abnormal chemical composition.

Effects resulting from over-secretion by an endocrine gland may be relieved or cured by the surgical removal of part of the hyperactive tissue. On the other hand the operative removal of too much tissue or the destruction of whole or part of an endocrine gland by disease gives rise to a 'deficiency syndrome' which is controlled, but never cured, by replacement therapy; that is to say by giving, usually by injection, extracts of the deficient gland obtained from animals or in some cases synthetic materials which have a similar effect. The amount of the replacement depends on the hormone content and production of the individual's own gland and the daily requirement also depends upon such factors as body weight, age, sex and activity.

The evolutionary history of the animal body shows that there has been a gradually increasing integration and control of the bodily functions. This is particularly true of the central nervous system which increases in size and complexity as we ascend the animal kingdom. Such co-ordination of function as exists in primitive animals without a nervous system is achieved by chemical means, and during the last fifty years it has become increasingly clear that a great deal of the behaviour even of higher animals and man is conditioned by the production and circulation of many hormones. Some of the most difficult problems in human physiology and in clinical medicine are concerned with the interrelationships between the actions of the various endocrine glands.

It is customary to restrict the term endocrine gland to the thyroid, parathyroid, pancreatic islet tissue, suprarenal cortex and medulla, the gonads and the pituitary body. The thymus and pineal are not included in this list since the evidence that either of these bodies produces an internal secretion is not conclusive. In the following pages we deal separately with all these glands except the gonads and the pituitary which are considered in Chapters 52 and 53 respectively. We end with a short account of the so-called local hormones.

THE THYROID GLAND

The thyroid gland develops in the floor of the embryonic mouth as a median diverticulum which grows downwards and backwards as a tubular duct with a bifurcated end that gives rise to the isthmus and part of the lateral lobes; the remainder of the lateral lobes is provided by cells derived from the fourth pharyngeal pouch. The connexion with the mouth, the thyroglossal duct, disappears early in development. The thyroid gland in the adult lies at the root of the neck overlying the larynx and trachea but it is sometimes prolonged into the thorax. Accessory thyroid tissue is occasionally found along the line of the original thyroglossal duct.

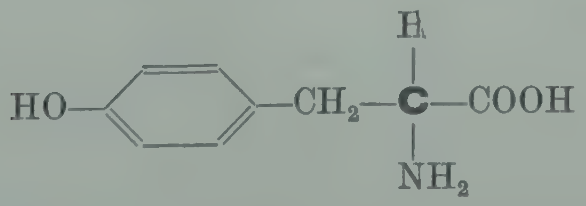
The thyroid gland, weighing about 25 g. in the healthy adult, consists of an enormous number of closed spherical vesicles, the walls of which are composed of cubical epithelium (Fig. 51, 3). The vesicles, which vary in size from 0·05 to 0·1 mm in diameter, are filled by a structureless semifluid colloid material containing the active principle of the gland. The thyroid gland has an abundant blood supply; some 4 to 6 ml. of blood per gram of tissue perfuse the normal gland per minute and the amount is much greater when the gland is overactive in disease. The sympathetic nerves which supply the gland do not appear to have any direct effect on the secretory processes but they exercise an important indirect control by regulating the blood flow through it.

The thyroid begins to secrete during fœtal life and at birth contains relatively large amounts of colloid material which gradually disappears during the first few weeks of extrauterine life. The gland is very active throughout infancy and childhood, reaching a peak of secretory activity at puberty, when the gland may actually appear to be enlarged and overactive. During adult life the thyroid is relatively less active, the secretory cells passing through phases of activity and rest until progressive involution begins about the fifth decade. It has been suggested that the gradual diminution in its activity in later life is one of the main factors responsible for the slower bodily and mental processes of the elderly. So far as is known at present thyroid activity is regulated by the thyroid-stimulating hormone (TSH) from the anterior pituitary (p. 1002). Because the factors regulating thyroid function are still obscure it is difficult to understand why the gland sometimes produces insufficient secretion giving the signs and symptoms of *hypothyroidism* or *myxœdema* or why it may become overactive producing the clinical state of *hyperthyroidism*.

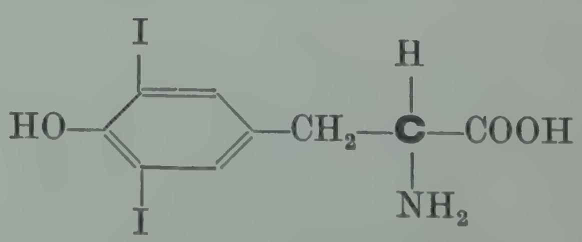
Desiccated thyroid gland is active when given by mouth and thyroid hormone is the only hormone as fully active when given by mouth as by injection. The potency of thyroid extracts is assayed biologically, the method most commonly employed being based upon the increase of metabolic rate in guinea-pigs or rabbits which follows ingestion or injection.

That iodine in firm organic combination was an important constituent of thyroid tissue was established by Baumann in 1895, shortly after the therapeutic value of extracts of the thyroid glands of animals in the treatment of hypothyroidism had been shown by Murray. This led to the preparation of more and more active extracts of thyroid tissue and culminated in the crystallization of thyroxine in 1919 by Kendall in the United States. About ten years later the method of isolation was greatly improved by Harington in Britain and the chemical constitution of thyroxine was determined by him in 1927. In the following year thyroxine was synthesized by Harington and Barger.

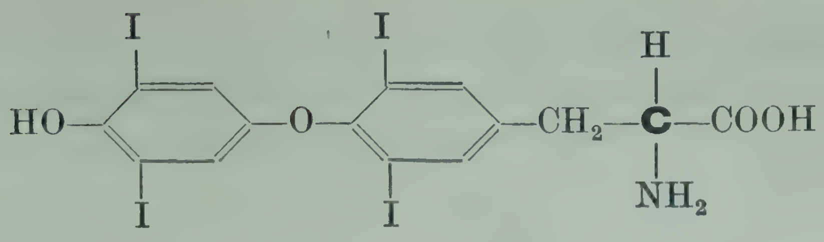
The thyroid gland contains about 0·06 per cent. of iodine, much more than any other tissue in the body, and this iodine is present both in the cells lining the vesicles and in the colloid material within them. If an equal volume of saturated ammonium sulphate is added to a saline extract of thyroid tissue an iodine-containing protein is found in the precipitate. This protein has a molecular weight of about 675,000; on hydrolysis it yields first thyroxine-polypeptide and then two characteristic amino-acids, diiodotyrosine and thyroxine.



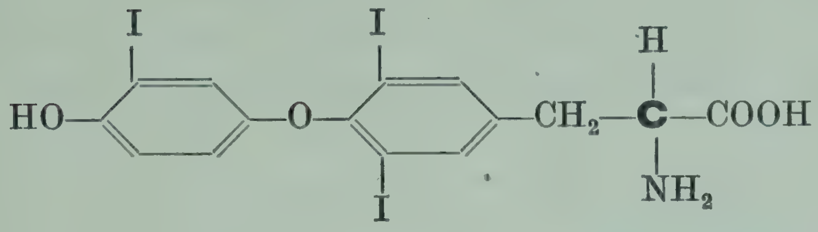
tyrosine



3:5-diiodotyrosine



thyroxine
(3 : 5 : 3′ : 5′-tetraiodothyronine)



3 : 5 : 3′-triiodothyronine

Both diiodotyrosine and thyroxine possess an asymmetric carbon atom and can therefore exist in optically active forms. Neither isomer of diiodotyrosine has any thyroid activity. L-Thyroxine is some three times more active than the D-isomer and, given by mouth or by injection, has the same quantitative effect as dried thyroid gland if both contain equal amounts of iodine. The activity of thyroid gland is proportional to its iodine, not to its thyroxine, content, and thyroid extracts are sometimes found to be more active physiologically than their content of thyroxine would allow. It has recently been found that triiodothyronine (thyroxine less one atom of iodine) is some four to eight times more potent than thyroxine itself. This compound is present both in the thyroid gland and in the blood and is in fact present in greater amounts than is thyroxine in the protein bound iodine fraction of the plasma. Triiodothyronine is now considered to be the hormone of the thyroid gland. The thyroid hormone appears to be manufactured by the epithelial cells of the thyroid vesicles from which it is discharged either directly into the blood vessels at the periphery of the vesicle or into the central space where it is stored in the form of 'colloid' until required.

Thyroxine is probably built up from the amino-acid tyrosine which becomes converted first into 3: 5-diiodotyrosine. Two molecules of 3: 5-diiodotyrosine may then be coupled together with the loss of one side chain. The points in favour of this view are as follows. In the first place the chemical structure of thyroxine is such as to make it probable that the hormone is derived from tyrosine. Secondly diiodotyrosine occurs as a normal constituent of the thyroid gland. In the third place the stereochemical configuration of natural thyroxine and natural tyrosine is such that the one could be derived from the other. Further support for this view was provided in 1939 by Ludwig and von Mutzenbecher, who showed that when casein was treated *in vitro* with iodine under suitable conditions minute amounts of thyroxine could be obtained. This suggested that the tyrosine of the casein was first converted to diiodotyrosine which was then transformed into thyroxine. Shortly afterwards Harington in London showed that diiodotyrosine could, in fact, be oxidized to thyroxine by hydrogen peroxide in faintly alkaline solution. Although the yields were small the presence of thyroxine was proved and it is likely that thyroxine is produced in a similar way when a protein is treated with iodine in a faintly alkaline medium which would favour the formation of hypoiodite to act as a suitable oxidizing agent. Such experiments, although indicating a probable route of formation, do not prove that thyroxine is synthesized by these steps in the thyroid gland itself.

This problem has been investigated with the help of radioactive iodine ^{131}I, chiefly by Chaikoff and his colleagues in California. When an animal is injected with minute amounts of radioactive iodine this is very quickly concentrated in the thyroid and two hours after the injection 10 to 15 per cent. of the total radioactive material can be detected in the thyroid gland although this organ is actually less than 0·01 per cent. of the body weight. The radioactive iodine is found in the thyroid gland in its three main fractions, namely inorganic iodine, diiodotyrosine and thyroxine. Soon after the injection the amount of radioactive iodine is small in the thyroxine fraction and large in the diiodotyrosine fraction, but later its concentration in the thyroxine as well as in the diiodotyrosine moiety increases. This is, of course, what would be expected on the assumption that diiodotyrosine is first formed and then converted to thyroxine. When sections of thyroid tissue are incubated *in vitro* with solutions containing radioactive iodine, the iodine is quickly incorporated into the tissue in inorganic combination, and the greater part of the radioactivity later comes to be associated with the diiodotyrosine fraction. While such an incorporation of iodine into diiodotyrosine can be brought about, both by the intact gland and by tissue slices taken from it, the process occurs much more slowly with minced gland tissue and does not occur at all with homogenized tissue. Thus the state of organization of the tissue is an important factor in the process. Furthermore, these experiments considered along with those of Harington described above suggest that an intracellular oxidizing enzyme system is involved in this synthesis. This idea is supported by the observation that the reaction does not occur in anaerobic conditions, nor does it occur even in the presence of oxygen when the slices are poisoned with substances such as cyanide. It has been shown, moreover, that when slices of thyroid tissue are incubated with radioactive iodine the uptake of iodine by the tissue is prevented by such substances as thiocyanate, whereas in the presence of thiourea and allied compounds the iodine is taken up by the slices but remains in inorganic form and is not incorporated into diiodotyrosine or thyroxine.

The isotope ^{131}I is of value in the diagnosis of thyroid disease in man. After a single dose by mouth (25 microcuries) in the form of sodium iodide about 2 per cent. is eliminated in the fæces and sweat. Approximately 15 per cent. is distributed throughout the body and the remaining 80 per cent. is either taken up by the thyroid or excreted in the urine. A larger proportion is taken up by an overactive than by a normal gland, and this difference may be detected either by a Geiger counter placed over the gland or by determining the amount of radioactivity in the urine passed during the twenty-four hours following administration (Fig. 51, 1). The uptake of radioactive iodine by thyroid tissue has been used with effect in the treatment of malignant disease of the gland. Here the concentration of radioactive material in the thyroid is built up to a level at which the growth of malignant cells is inhibited.

The injection of ^{131}I into animals has been used to study the factors controlling the output of hormone from the thyroid gland. If a single injection of TSH is given there is, after a latent period of about an hour, a considerable reduction in the count obtained over the neck which indicates, of course, an increase in the output of ^{131}I, that is an increase in output of hormone. For a short time after this the output of ^{131}I is completely inhibited owing to the suppression of the animal's own TSH secretion by the raised level of thyroid hormone in the blood. An injection of thyroxin can also prevent the output of ^{131}I completely for a considerable time. Exposure of rabbits to moderate cold increases the output of ^{131}I from the thyroid gland but intense cold, emotional stress, or physical trauma reduce the output of hormone. Injection of cortisone or ACTH inhibits the output of ^{131}I from the thyroid, presumably mainly by suppressing the secretion of TSH. Hypophysectomy, as would be expected, causes a prompt and permanent decrease in the rate of release of ^{131}I from the thyroid.

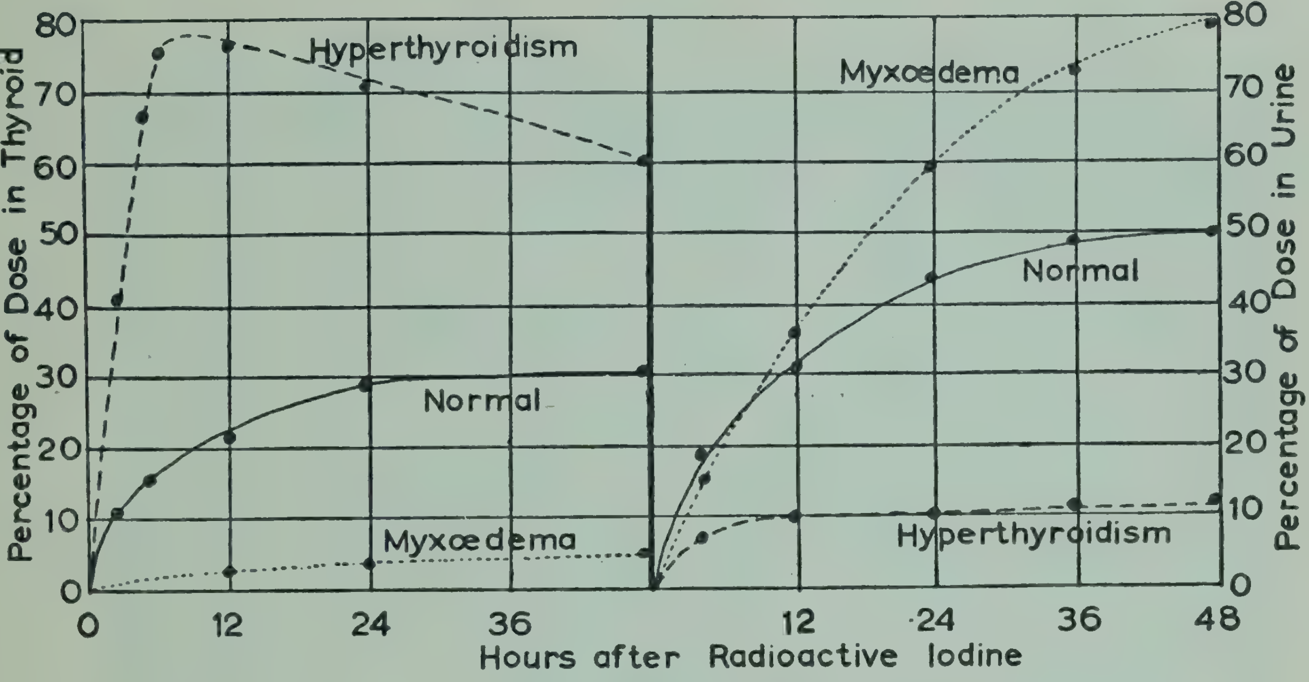


FIG. 51, 1

Uptake of ^{131}I by the thyroid gland and urinary excretion of ^{131}I in a normal person and in typical cases of hyperthyroidism and myxœdema (C. P. Stewart & D. M. Dunlop (1954), *Clinical Chemistry in Practical Medicine*. 4th Ed. Edinburgh: E. & S. Livingstone.

Iodine is found in the blood in two forms, inorganic iodide and protein-bound iodine, the latter being loosely attached to serum albumin. The protein-bound iodine includes thyroxin, triiodothyronine and diiodotyrosine and its concentration in the plasma is, therefore, an index of the amount of hormone in the blood and of the secretory activity of the thyroid gland. The normal level is some 3 to 20 μg. per 100 ml. In myxœdema it is about 1 μg. per 100 ml., while in hyperthyroidism it is over 20 μg.

Measurement of the concentration of protein-bound iodine in the plasma is difficult, but the protein-bound radio-active iodine after a dose of ^{131}I can quickly and easily be determined. Such measurements form the basis of a useful test for thyroid activity (Fig. 51, 2).

The thyroid hormone plays an important part in controlling the rate of oxidative processes in the cells of the body; the oxygen uptake *in vitro* of tissues from thyroidectomized animals is subnormal but it can be increased to very high levels by the addition of thyroxine. The activity of the thyroid gland in an individual or the effects of giving him extracts of thyroid gland can be assessed by periodical measurements of the basal metabolic rate.

Hypothyroidism.—Thyroidectomy always produces a marked fall in the metabolic rate, with subnormal temperature, slow pulse and respiration and mental apathy. The effects of this profound metabolic disturbance are somewhat different in the immature and adult animal. Hypothyroidism in the human infant is described as *cretinism* (Fig. 51, 4). Although the condition may not be recognizable until three or four months after birth, it is probable that the thyroid gland in these cases has never functioned. The child is small, mentally defective, with coarse, scanty hair and a thick, yellowish scaly skin. Hypothyroidism occurring in the adult is called *myxœdema* (Fig. 51, 5). There is gradual retardation of mental and bodily processes, the basal metabolic rate may fall to 40 per cent. below normal and the patient feels cold. Speech and cerebration are slow and sexual function is depressed. There is considerable increase in body weight. The hair is thin on the scalp and eyebrows, the face and eyelids become puffy, the tongue swells and the skin becomes rough and coarse. The puffiness of the hands and face and the thickening of the skin are due in part to excessive fluid in both extra- and intra-cellular compartments, and in part to the deposit in the tissues of a semifluid material containing over 13 per cent. of protein.

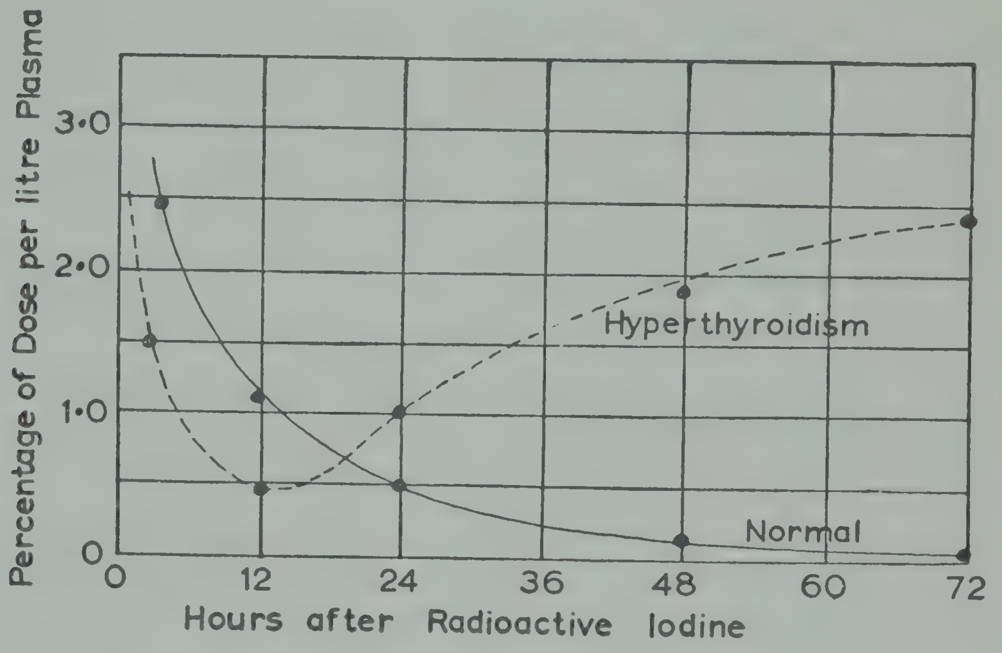


FIG. 51, 2

The plasma ^{131}I following a test dose in a normal person and in a patient suffering from moderate hyperthyroidism. In the patient with hyperthyroidism the reappearance of radioactivity in the blood is the result of discharge of hormone containing ^{131}I from the active gland. At the end of 48 hr after the dose of isotope the hyperthyroid patient has therefore more protein-bound radioactivity than the normal (C. P. Stewart & D. M. Dunlop (1954), *Clinical Chemistry in Practical Medicine*. 4th Ed. Edinburgh: E. & S. Livingstone.

Murray in 1891 was the first to relieve myxœdema by the administration of a glycerine extract of sheeps' thyroids. Thyroid substance is active by mouth, but some twelve to twenty-four hours must elapse before increased oxygen consumption can be demonstrated and the maximal effect is not obtained for about seven days (Fig. 51, 6). This is in contrast to its immediate effect on the oxygen utilization of tissue slices. The reason for

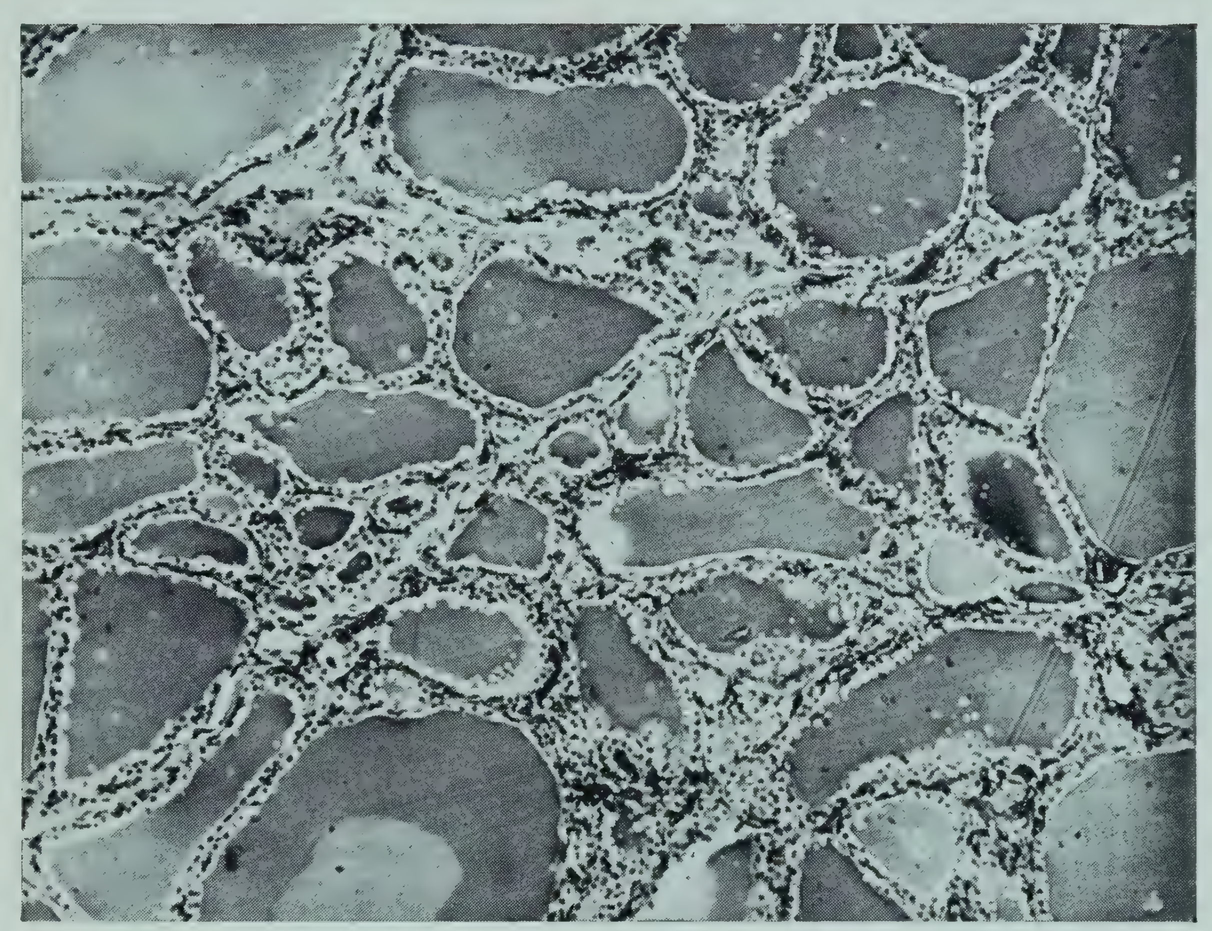


Fig. 51, 3

Photomicrograph (× 100) of human thyroid tissue showing vesicles containing colloid lined by cubical epithelium.

(*By courtesy of J. W. Orr.*)

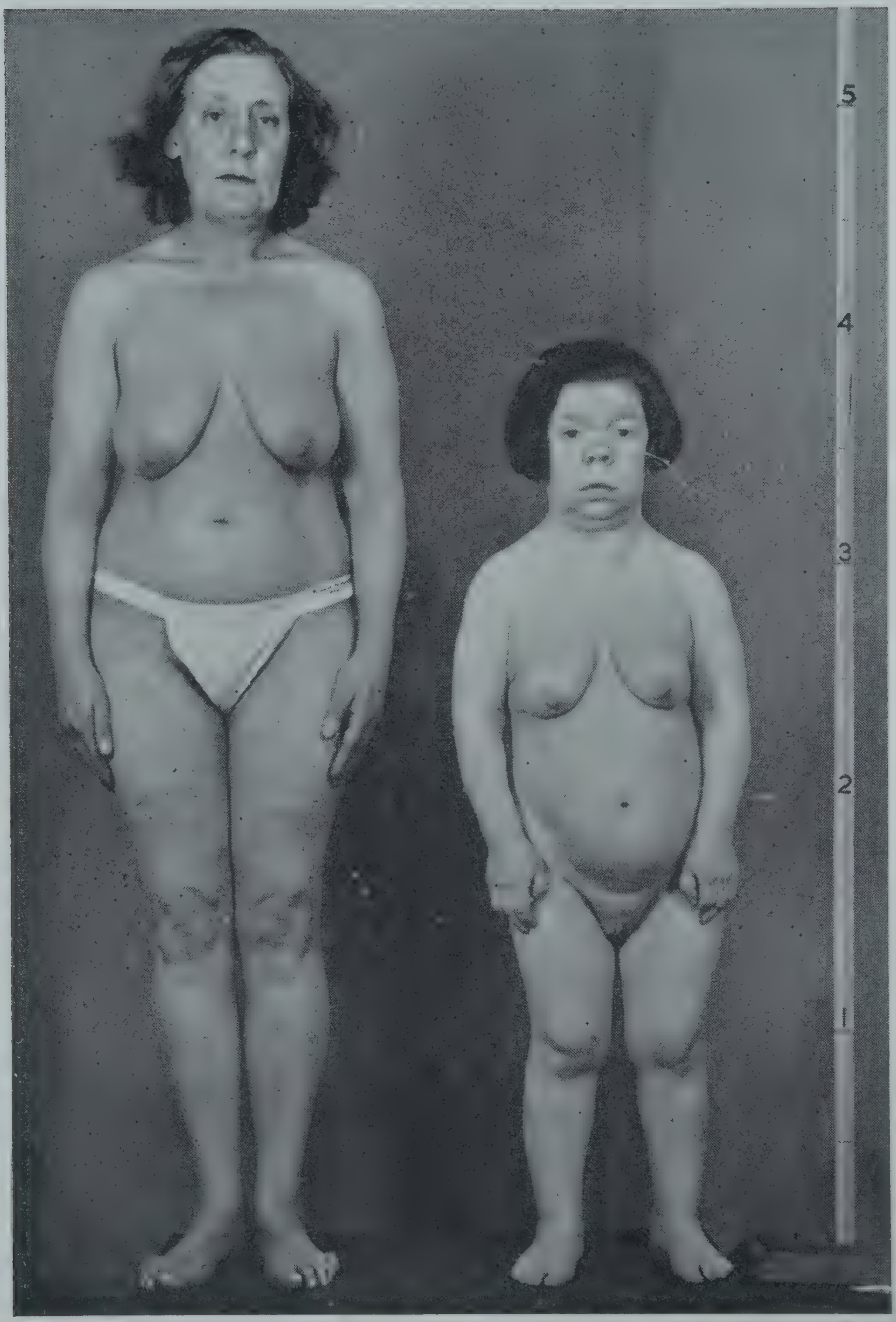
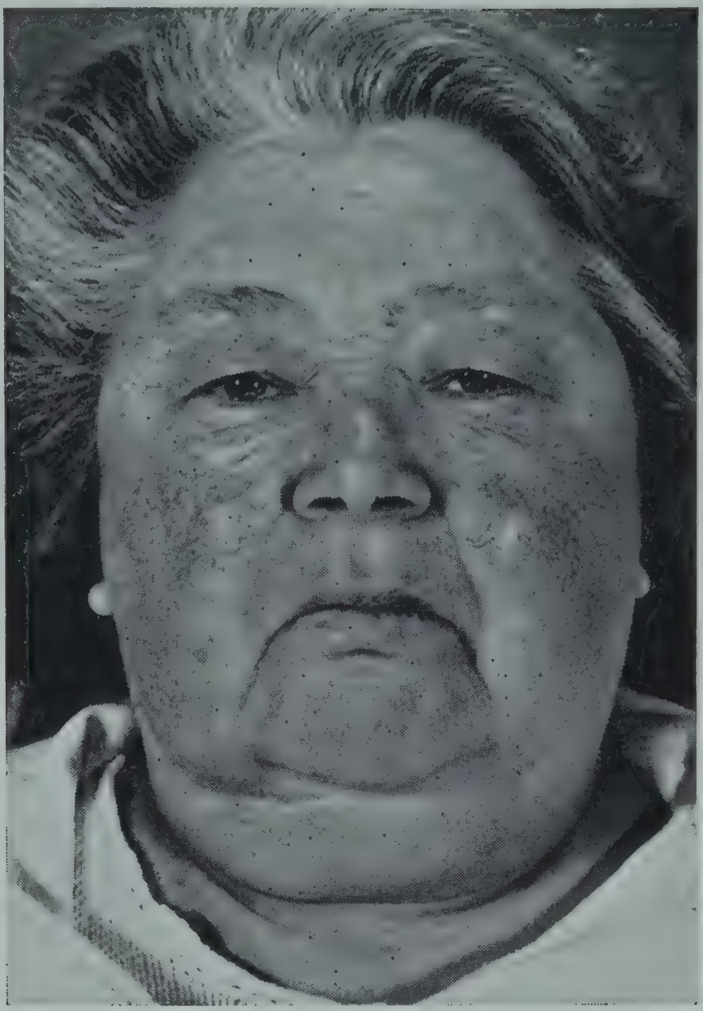


FIG. 51, 4

On the right a cretin, aged 38 years, and on the left her normal sister, aged 48. The mental age of the cretin is about 4 years. She has a deep, croaking voice, coarse dry skin and hair.

(By courtesy of A. C. Crooke.)



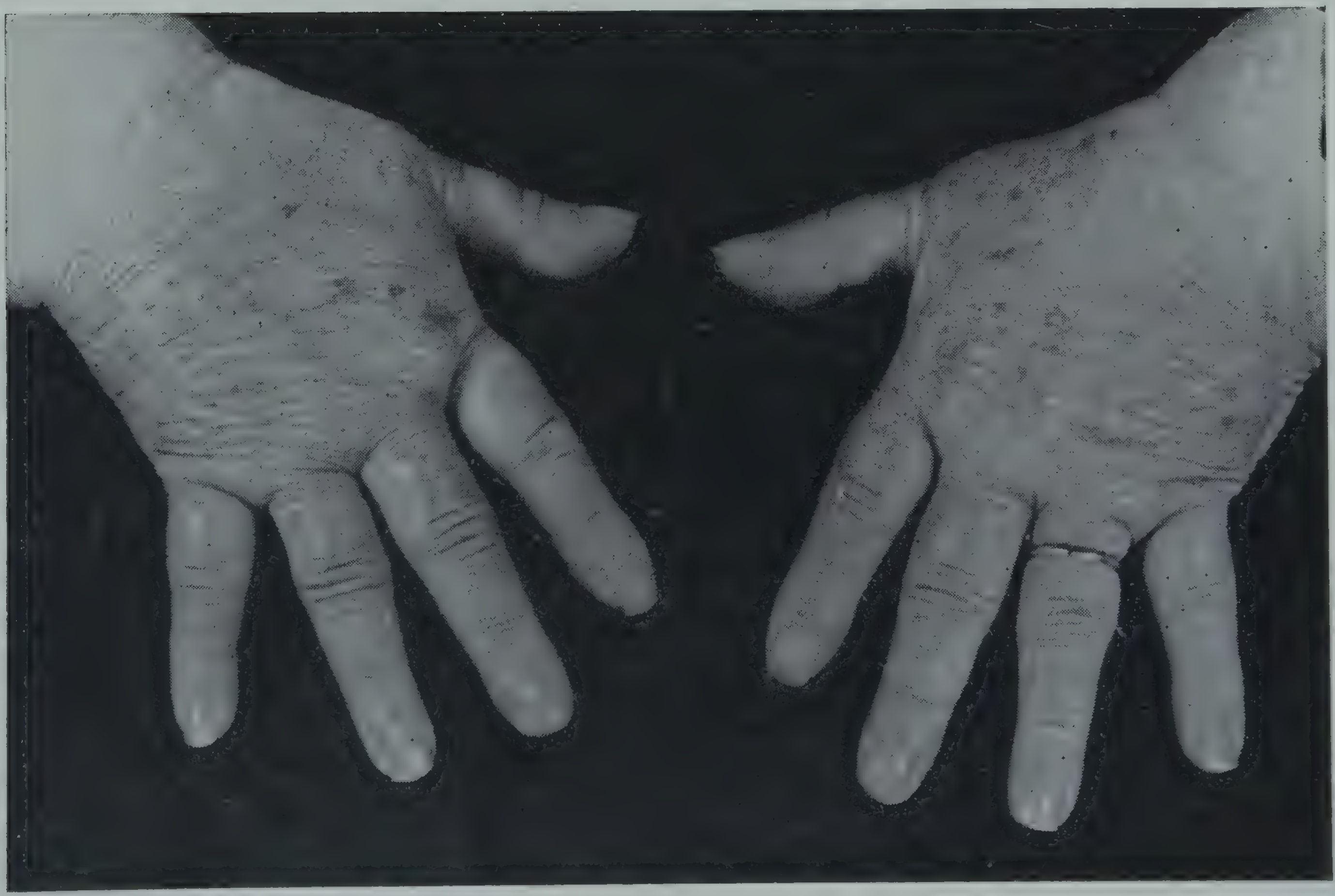


FIG. 51, 5

Photographs of the face and hands of a woman, aged 69, suffering from myxœdema. Note the dull expression, scanty eyebrows and dilated vessels on the cheeks. The face and hands are puffy, and the skin of the hands is dry and wrinkled.

(By courtesy of W. T. Cooke.)

the delayed action *in vivo* is quite unknown. Hypothyroidism can be relieved by the administration of desiccated thyroid gland or of synthetic L: thyroxine or of tri-iodo-thyronine. In animals the gland can be replaced by a graft.

Hyperthyroidism.—The administration of an excessive amount of thyroid hormone leads to an increase of the metabolic rate, the body temperature is elevated and the skin is warm and moist as the body attempts to get rid of the extra heat. The patient loses weight as the result of increased catabolism of tissue protein and oxidation of the stored fat. Similar effects are produced by an increase in the secretory activity of the thyroid gland which is often accompanied by an enlargement of the gland (a *goitre*). The patient has marked nervous excitability and a fine involuntary tremor (Fig. 51, 7). He is breathless on exertion, has a rapid heart rate, a warm flushed skin and an anxious staring expression. Protrusion of the eyeballs (exophthalmos) may also be present and this may persist long after the other features have been abolished by treatment. In animals exophthalmos has been produced by administration of pituitary TSH but not by thyroid hormone.

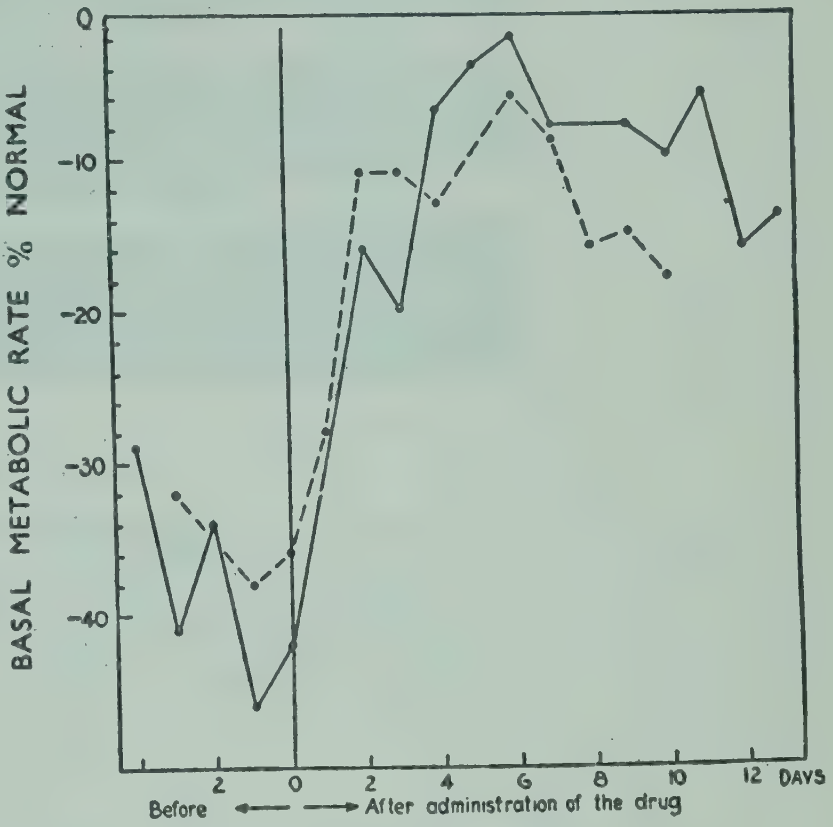


FIG. 51, 6

Comparison of the effects of a single dose of 13 mg. of thyroxine polypeptide containing 6·5 mg. of iodine given to one case of myxœdema intravenously (solid line) and to another case by mouth (broken line). Note the delay before the maximal effect is obtained on metabolism. The effect on the basal metabolic rate is very similar in the two cases (W. T. Salter, J. Lerman, and J. H. Means (1933) from J. H. Means (1948), *The Thyroid and its Diseases.* 2nd ed. Philadelphia: Lippincott).

The effects of excessive thyroid activity can be overcome by the surgical removal of part of the overactive gland (thyroidectomy) but a more normal histological appearance of the gland may be restored by giving iodine which greatly reduces the output of hormone (Fig. 51, 8). The amount of iodine required to produce a maximal effect is of the order of 6 mg. per day, and there is little to be achieved by giving doses greatly in excess of this quantity. This dose should be compared with the daily intake adequate for the prophylaxis of endemic goitre (p. 928), namely about 0·075 mg., and with the average daily intake of healthy people, about 0·1 to 0·2 mg. In spite of continued administration of iodine, however, the metabolic rate after two or three weeks tends to rise and the signs and symptoms of hyperthyroidism reappear.

In 1941 Mackenzie, Mackenzie and McCollum observed that sulphaguanidine administered to rats over a long period of time caused a remarkable enlargement of the thyroid gland. Later it was found that a similar effect could be more readily induced by the administration of thiourea and thiouracil, and these substances are now known as *goitrogens* or *antithyroid substances.* Their administration to rats results in

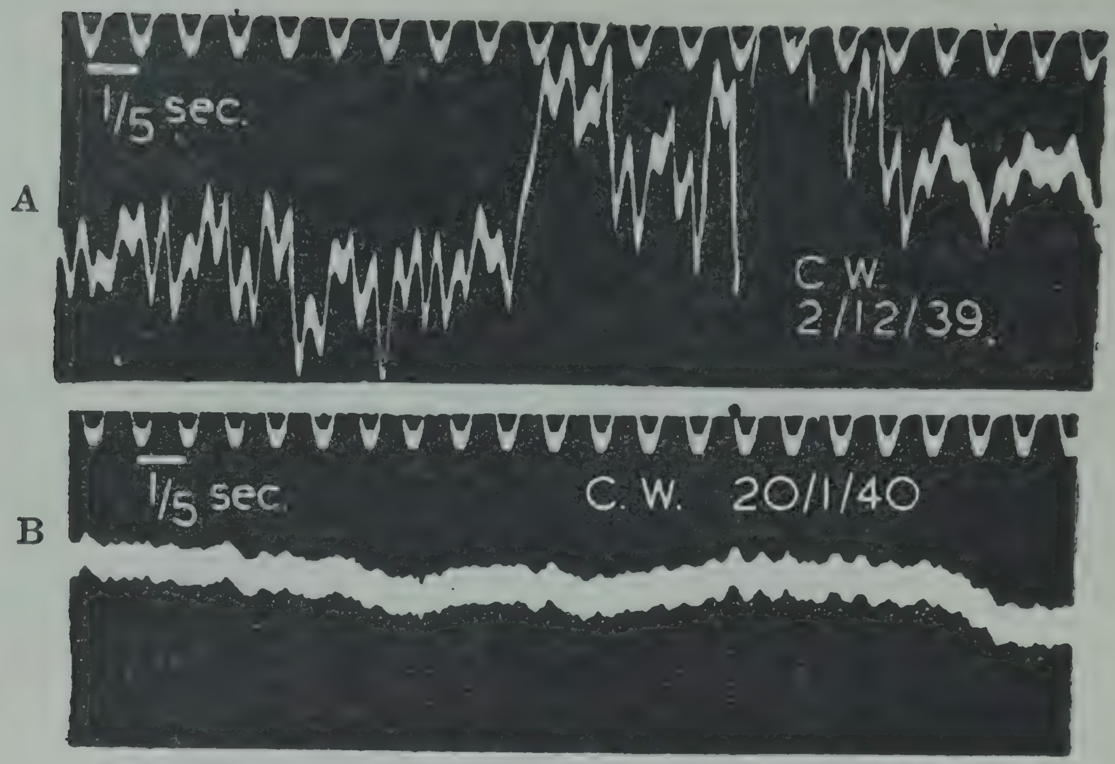


Fig. 51, 7

Muscular tremor as shown by the magnified movements of the tip of the extended index finger of a patient, aged 20, with hyperthyroidism before operation (upper tracing), and 45 days after partial removal of the thyroid gland (lower tracing) (S. Lazarus & G. H. Bell (1943), *Glasg. med. J.* **140,** 77).

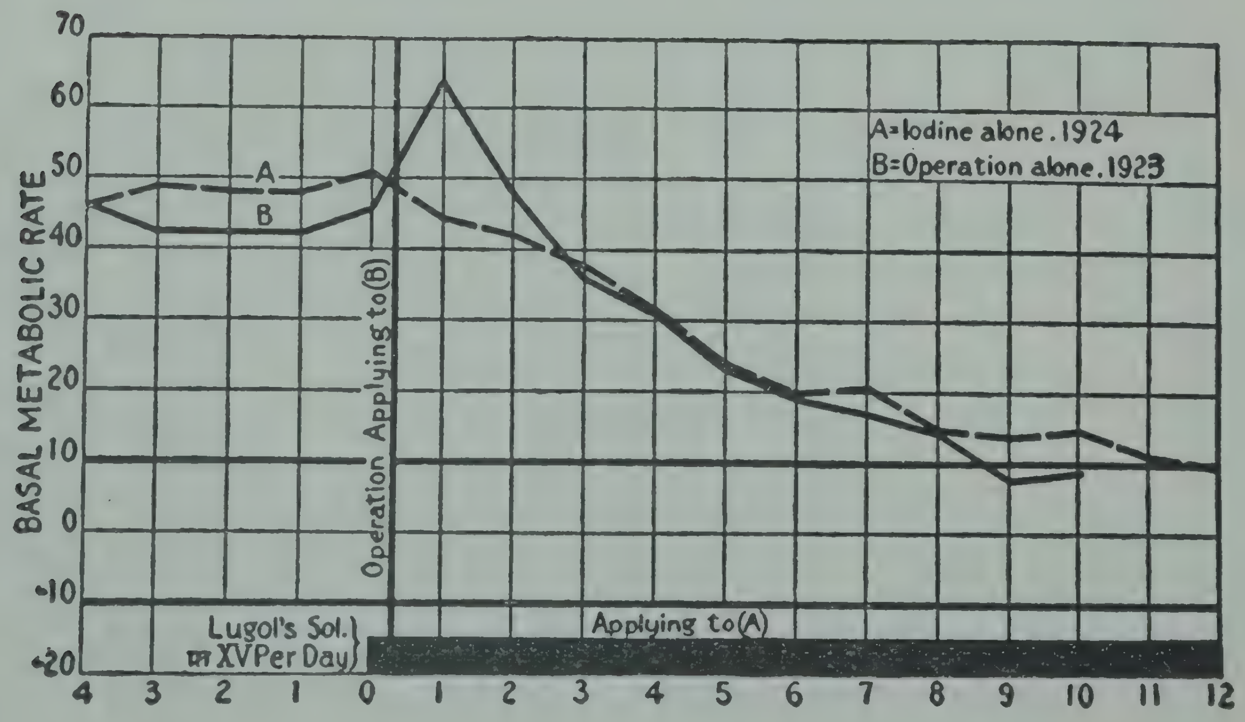


Fig. 51, 8

The effect of iodine compared with that of partial removal of the thyroid gland in the untreated hyperthyroid patient. The solid line represents the average curve obtained by Segall & Means (1924) in uniodinized patients; the broken line the average curve for iodine response obtained by Starr, Segall & Means (1924). Lugol's solution containing 5 per cent. iodine and 10 per cent. potassium iodide was administered daily in a dose of 15 minims (=1 ml.). Except for the temporary increase of B.M.R. after operation, the curves are remarkably similar (J. H. Means (1948), *The Thyroid and its Diseases.* 2nd ed. Philadelphia: Lippincott).

proliferation of the epithelial cells of the vesicles of the thyroid gland with complete loss of colloid, the histological picture similar to that found in severe untreated cases of hyperthyroidism. At the same time, although the gland is apparently overactive as judged by histological appearances, it is actually secreting less hormone than normal and the basal metabolic rate falls. The cellular proliferation of the gland is unaffected by iodine but it is prevented by the administration of thyroid hormone, and animals treated with thiourea or a related compound show a normal physiological reaction to the administration of thyroxine. It has also been observed that the administration of goitrogens to animals is followed by a sudden sharp drop in the concentration of iodine in the thyroid gland. When potassium iodide is administered to such animals the iodide becomes concentrated in the thyroid gland with remarkable speed but it remains in the gland in the form of inorganic iodide and is not converted into the hormone. All these observations have led to the view that the effect of goitrogens is to inhibit the synthesis of the hormone in the thyroid gland, and such substances have been employed with considerable success in the treatment of hyperthyroidism in patients.

Thyroid enlargement occurs in intact animals after administration of goitrogens but not in hypophysectomized animals. The explanation usually given is that, when the synthesis of hormone is inhibited in intact animals by the administration of goitrogens, the low concentration of hormone in the blood causes an increased output of TSH from the anterior pituitary, and this in turn brings about hyperplasia of the thyroid.

Endemic Goitre.—Since iodine is an essential part of the active principle of the thyroid gland, it is to be expected that a diet deficient in

TABLE 51, 9

Incidence of enlargement of the thyroid gland in children, aged 11 to 15 years, from four areas in which the drinking water varied greatly in iodine content

Area	Iodine content of water (μg. per litre)	No. of children examined	Total no. of visible thyroid glands	Percentage incidence of visible thyroid glands
Okehampton, Devon	1·1 (very low)	298	78	26·2
North Oxfordshire .	2·0 (low)	451	81	18·0
Windsor, Berks .	10·1 (high)	461	31	6·7
Maldon, Essex .	50·2 (very high)	527	13	2·5

From Table I of *M.R.C. Memorandum* (1948), No. 18. London: H.M.S.O.

iodine might lead to difficulties in the manufacture of the specific hormone. These difficulties are shown by a swelling of the thyroid gland. In districts where the supply of iodine in the water is deficient a large part of the population may be goitrous, hence the name endemic goitre. The children of goitrous mothers are often cretins. There is an inverse relationship between the intake of iodine and the incidence of goitre (Table 51, 9), but other deficiencies such as that of vitamin A may play a part in producing goitre. The first large-scale trials of the effect of giving small amounts of iodine to a population in which goitre was endemic were made in Ohio in 1916. The results were quite remarkable as the following figures show; only 0·2 per cent. of non-goitrous girls who took iodine over the four years of the experiment showed an enlarged thyroid gland, whereas 27·6 per cent. of those not taking iodine became goitrous. Since then, iodine has been administered to large populations either as sodium or potassium iodide or as potassium iodate or as iodized fat; the usual method is to add a small amount of iodide to the table salt. As a result of this measure the incidence of goitre in certain districts has fallen by about 90 per cent. in the general population and the incidence of goitre in the new-born by the same amount. Therefore a Sub-Committee of the Medical Research Council has recommended (1944) that all table salt sold in Britain should be iodized by the addition of one part of potassium iodide to 100,000 parts of sodium chloride, or of one part of potassium iodide to 40,000 parts of packeted table salt. These measures would provide the daily iodine requirement of 200 μg.

THE PARATHYROID GLANDS

The parathyroid glands develop from the endodermal lining of the third and fourth pharyngeal pouches of the embryo. They lie in the neck immediately adjacent to the posterior surface of the thyroid gland with which, however, they have no physiological relationship. They consist as a rule of four oval bodies about 6 mm long but they are variable in number, size and position; accessory parathyroid tissue is not uncommon lower in the neck or even in the thorax. The parathyroids are composed of epithelial cells, arranged for the most part in closely packed masses, separated by irregular bands of connective tissue richly supplied with blood vessels. Here and there however in human glands, an acinar arrangement is evident. The cells forming the greater part of the glands are small with relatively large nuclei and pale non-granular cytoplasm and are the active secretory tissue; the function of the large oxyphil cells containing eosinophilic granules is unknown (Fig. 51, 10).

In contrast to the rich vascular supply the nervous connexions are scanty and probably vasomotor. Secretory nerve fibres have never been demonstrated, and transplanted glands can function in the absence of all nervous attachments. Parathyroid activity is not controlled by a trophic hormone from the anterior pituitary but probably by alterations in the concentration of calcium in the blood.

In 1924 Collip prepared the first active extracts of parathyroid tissue; more potent preparations have now been obtained by removing most of the

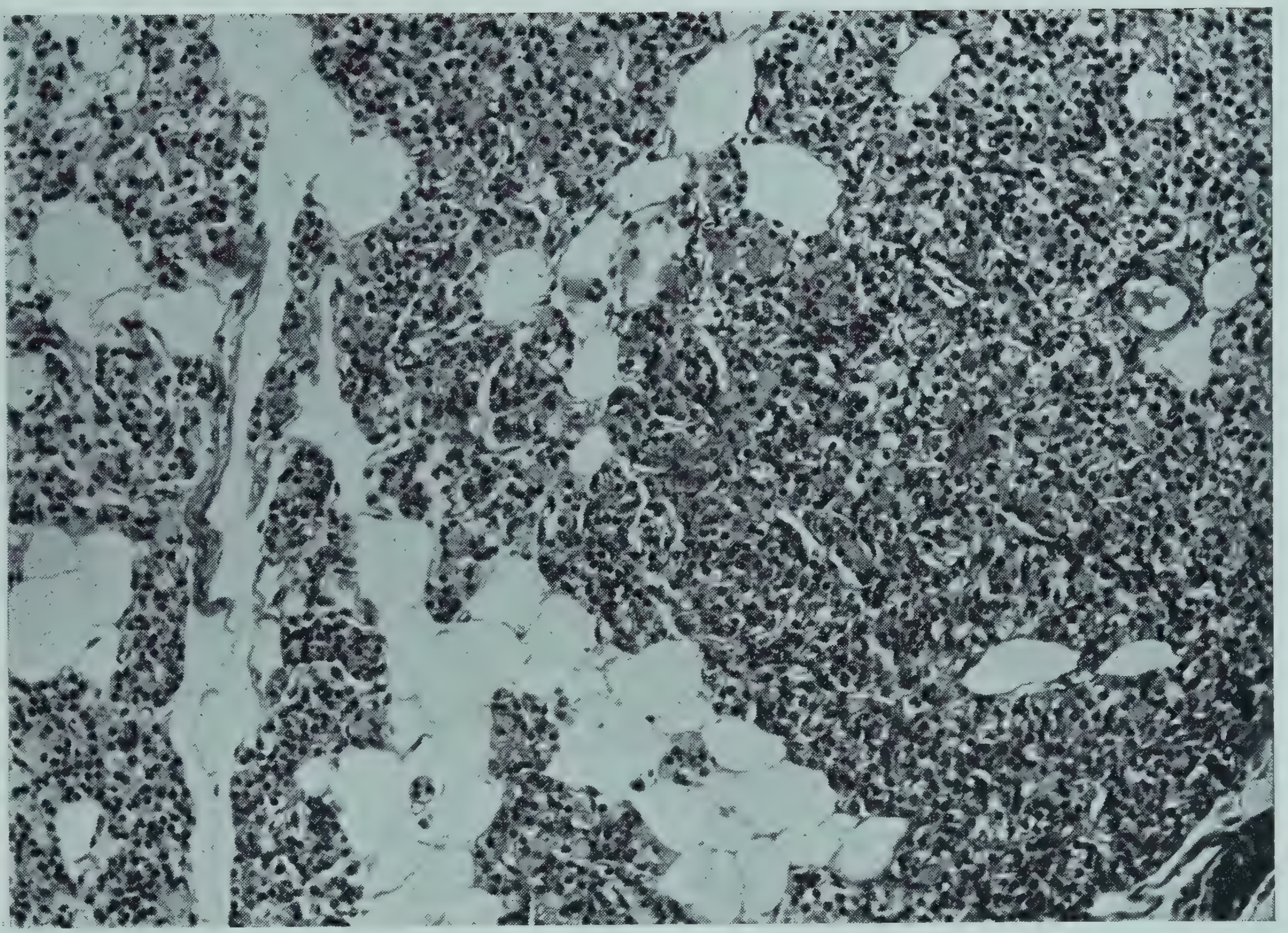


Fig. 51, 10

Photomicrograph (× 146) of human parathyroid tissue showing the densely packed secretory cells and occasional large oxyphil cells.

(*By courtesy of J. W. Orr.*)

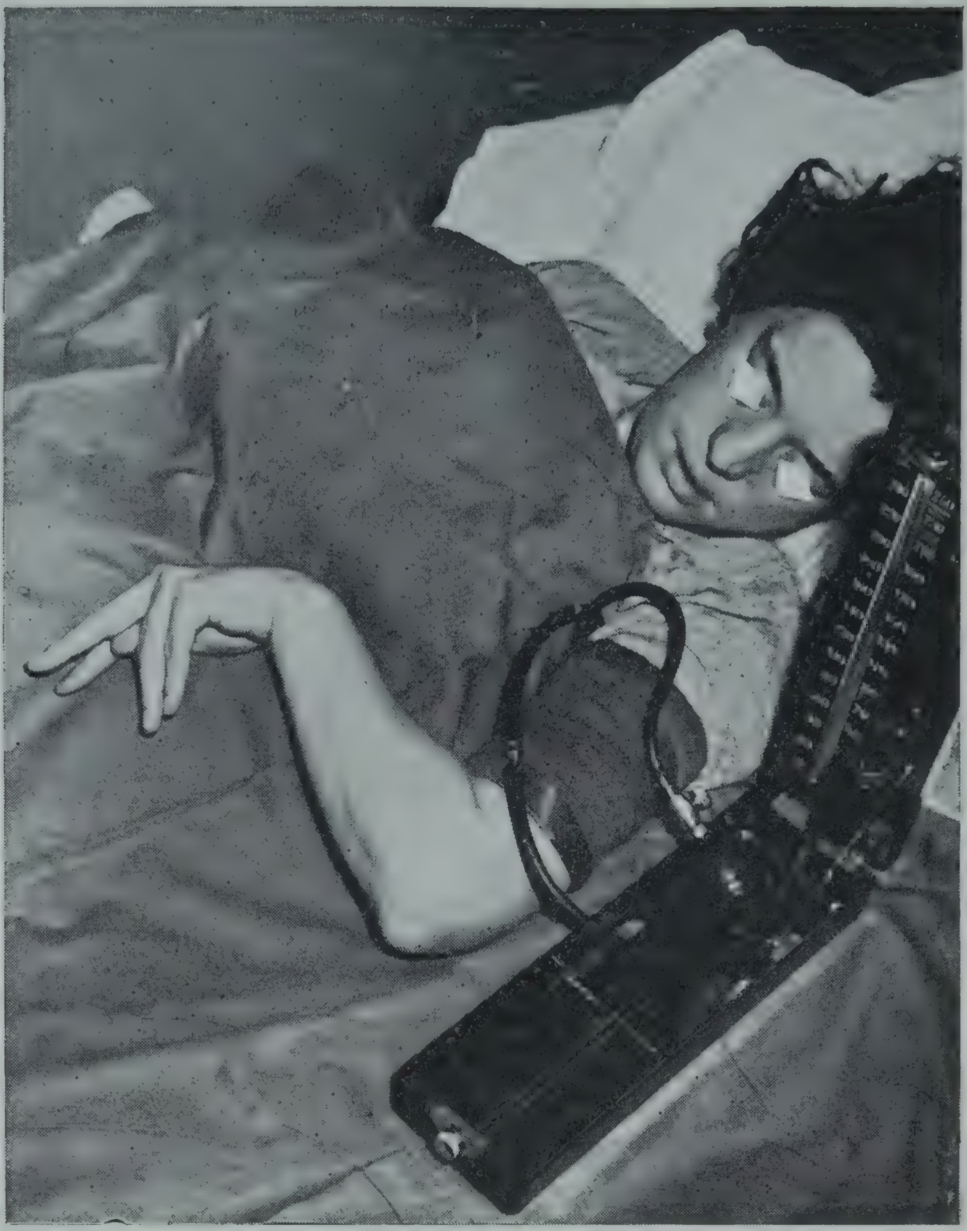


Fig. 51, 11, A

Main d'accoucheur induced by arrest of the circulation in the forearm in a patient with parathyroid tetany due to the accidental extirpation of the parathyroid glands during thyroidectomy for Graves's disease (parathyropriva tetany).

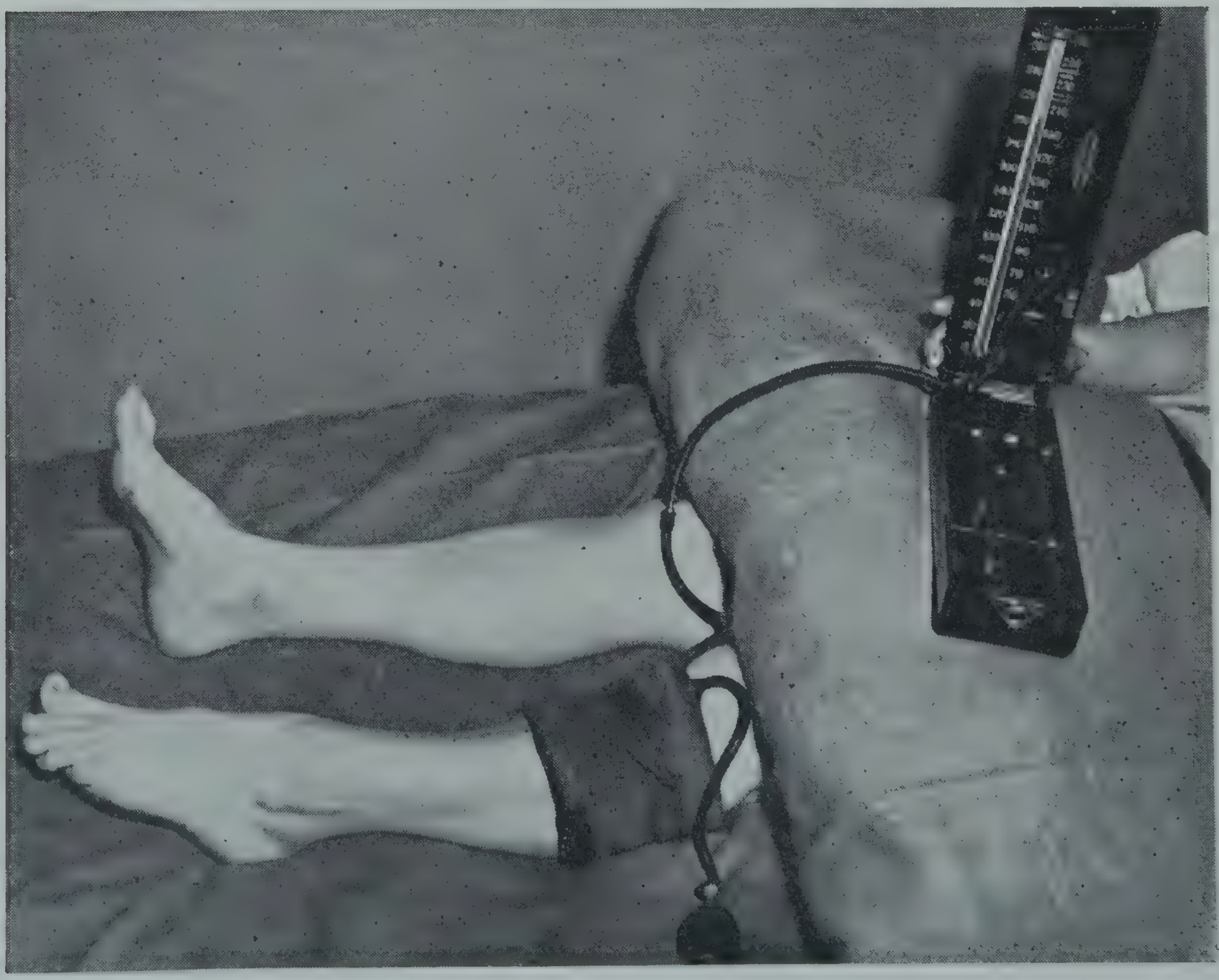


Fig. 51, 11, B

Pedal spasm induced in the same patient.

protein from an acid extract and then adsorbing the hormone on benzoic acid. The active material, called PTH (parathyroid hormone), behaves like a protein with a molecular weight about 650,000. It is readily destroyed by pepsin and trypsin and is therefore inactive when given by mouth.

A unit of parathyroid activity has been defined as one-hundredth of the amount of hormone required to increase the serum calcium of a dog weighing between 10 and 12 kg. by 1 mg. per 100 ml. within 16 hours. This method of standardizing extracts is never accurate and is not a satisfactory means of assessing the effects of parathyroid extracts in man.

Collip in 1925 first demonstrated that PTH elevated the serum calcium concentration in both normal and parathyroidectomized dogs, and it regularly does so also in man. Whether this phenomenon is primarily the result of an action on the kidney or due to resorption of calcium from the bones cannot yet be decided.

When an injection of PTH is given to a patient with hypoparathyroidism the first demonstrable effect is an elevation in the urinary elimination of phosphorus (Fig. 51, 12). This is followed by a fall in the concentration of phosphate in the serum and subsequently by a rise in the serum calcium. Later, there is an increased elimination of calcium in the urine. Albright believes that the following sequence of events occurs when PTH is administered:

(1) increased excretion of phosphorus in the urine.
(2) decreased concentration of phosphate in the serum.
(3) resorption of phosphorus and calcium from bones.
(4) elevation of the concentration of calcium in the serum.
(5) increased urinary elimination of calcium.
(6) further resorption of phosphorus and calcium from bones with consequent weakening of the skeleton.

This description is consistent with the view that PTH inhibits the reabsorption of phosphate by the renal tubules.

On the other hand the injection of PTH into nephrectomized dogs produces an elevation of serum calcium concentration within eight hours, and there is much evidence that PTH directly stimulates osteoclastic activity, leading to resorption of bone salts. The changes in the concentration of calcium and phosphorus in the serum and urine may then be secondary phenomena.

The relation between removal of the parathyroids and *tetany* was established by Gley in 1892. The outstanding feature of tetany is neuromuscular irritability which manifests itself first as a generalized hypertonicity of muscles and then as fibrillary twitchings of muscle fasciculi, leading finally, especially in children, to generalized clonic movements. The muscle hypertonia produces the characteristic attitude of the hand in tetany, the so-called *main d'accoucheur*. Simultaneously the feet are held firmly flexed at the ankles with all the toes plantar-flexed. The condition, called *carpopedal spasm* (Fig. 51, 11), is frequently associated in infants with spasm of the glottis (*laryngysmus stridulus*), which may be severe enough to cause alarming cyanosis. In the early stages of tetany sensory phenomena are often present, such as widespread tingling feelings and sensations of heat and flushing (paræsthesiæ). The neuromuscular hyperexcitability can be demonstrated before tetany occurs (*latent tetany*).

Tapping over the facial nerve in front of the ear produces twitching of the facial muscles (Chvostek's sign), and the motor nerves can be shown to be unduly excitable to galvanic stimulation. Carpal spasm can be induced by inflating a blood pressure cuff round the upper arm to a pressure exceeding the systolic blood pressure and maintaining the occlusion for three minutes (Trousseau's test) (Fig. 51, 11); ischæmia of nerve trunks itself increases their excitability and thus reinforces the effect of a low concentration of calcium in the serum.

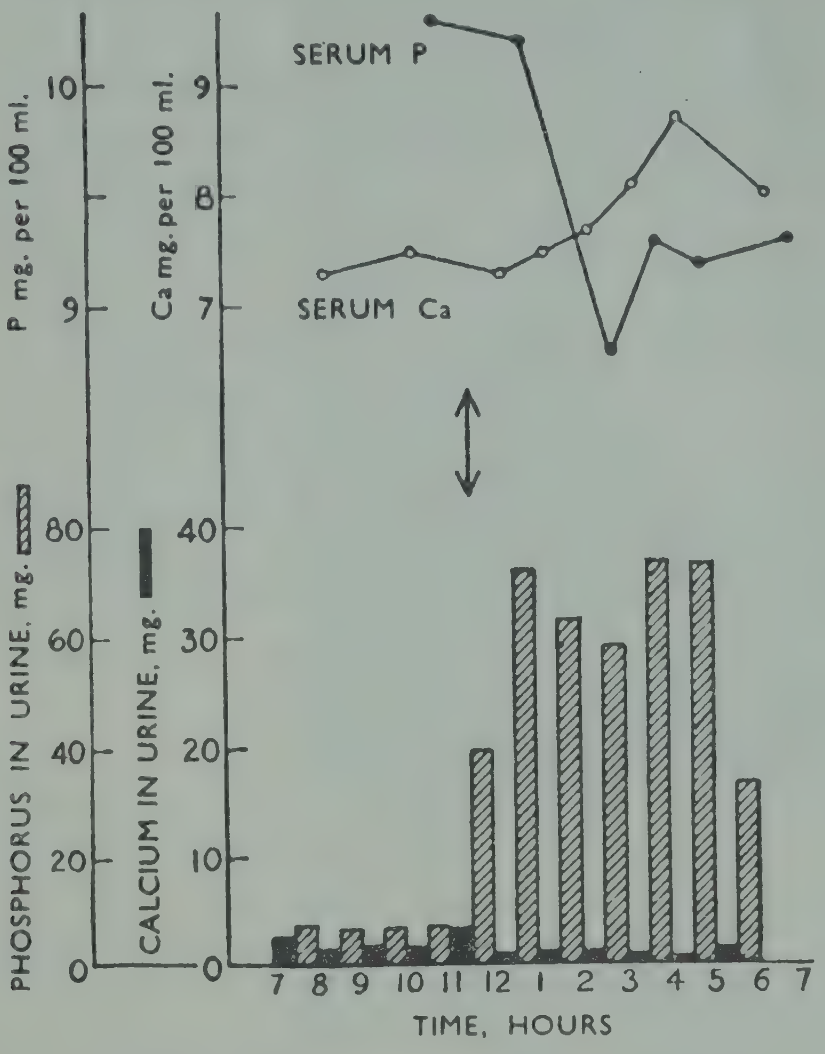


FIG. 51, 12

Effect of parathyroid extract on serum levels of calcium and phosphorus and urinary calcium and phosphorus excretion in a patient with idiopathic hypoparathyroidism; 75 units of parathormone were injected at the arrow (F. Albright & E. C. Reifenstein (1948), *Parathyroid Glands and Metabolic Bone Disease.* Baltimore: The Williams & Wilkins Co.).

The relationship between calcium and neuromuscular hyperexcitability was demonstrated by Loeb who showed that excitability of frog muscle was increased by sodium and potassium ions and decreased by calcium and magnesium ions. However, in tetany the concentration of calcium in the tissues is not abnormal, and the severity of tetany is not closely correlated with the extent of any reduction in the level of the serum calcium concentration. The level of blood calcium must, therefore, be only one of the factors involved in producing the increased irritability of muscles and nerves.

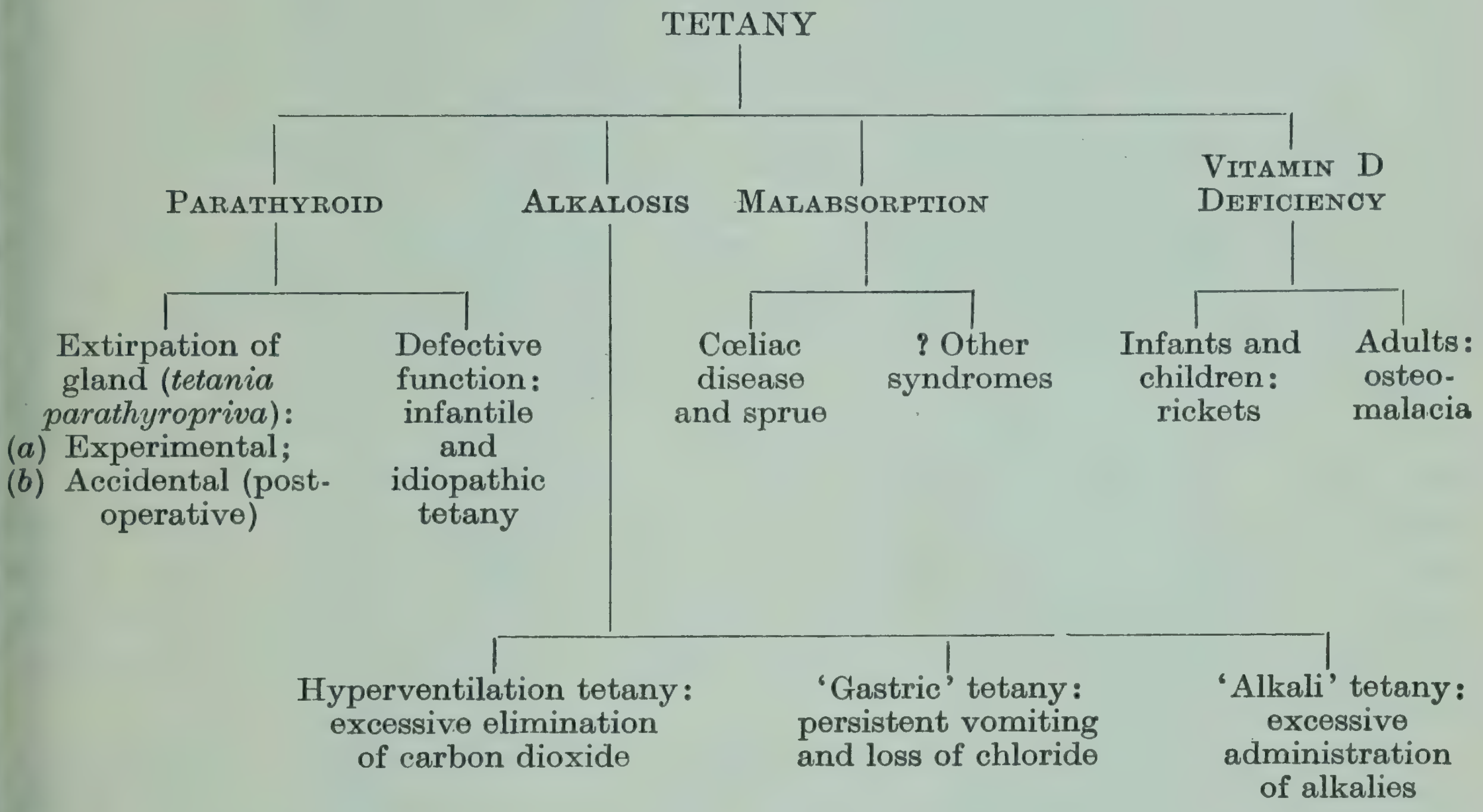
Of the normal 9 to 11 mg. of calcium per 100 ml. of normal serum 42 to 58 per cent. is present in a dialysable (diffusible) form and of this probably over 80 per cent. is ionized, normally about 4·5 to 6·5 mg. per 100 ml. (p. 101). A fall in the concentration of ionic calcium has been held to be responsible for the neuromuscular irritability which culminates in tetany. This hypothesis is, however, hard to establish because of the difficulty of assaying the serum concentration of ionized calcium. Tetany is abolished by curare and probably depends on the integrity of spinal reflex arcs. It is in fact clear that a number of factors is involved in the production of tetany, for example low serum magnesium, low plasma protein, hyperphosphatæmia and alkalosis.

The causes of tetany met with clinically are summarized in Table 51, 13. Treatment should aim at restoring the blood concentration of ionized calcium. The diet should therefore be low in phosphorus but rich in calcium and should be supplemented with a soluble calcium salt and with vitamin D. Intravenous injection of calcium gluconate usually arrests an attack of tetany. The administration of parathyroid hormone may be required to control the blood calcium concentration in patients from whom the glands have been inadvertently removed.

Actively secreting tumours of the parathyroid glands produce excessive amounts of PTH, and the resulting disturbance of phosphorus and calcium metabolism leads to withdrawal of large amounts of these elements from the bones which therefore become weak and deformed and liable to fracture (Fig. 51, 14). The serum calcium may be as high as 16 mg. per 100 ml. and the urinary excretion may be even greater than the fœcal. At the same time the serum phosphorus falls to 1 or 2 mg. per 100 ml. As in all cases where bone is being reabsorbed the serum alkaline phosphatase is raised. The increased concentration of calcium in the urine may lead to the formation of renal stones. Removal of the tumour is often followed by cure but may result in hypoparathyroidism.

TABLE 51, 13

Causes of Tetany


TETANY
PARATHYROID
ALKALOSIS
MALABSORPTION
VITAMIN D DEFICIENCY
Extirpation of gland (*tetania parathyropriva*): (*a*) Experimental; (*b*) Accidental (post-operative)
Defective function: infantile and idiopathic tetany
Cœliac disease and sprue
? Other syndromes
Infants and children: rickets
Adults: osteomalacia
Hyperventilation tetany: excessive elimination of carbon dioxide
'Gastric' tetany: persistent vomiting and loss of chloride
'Alkali' tetany: excessive administration of alkalies

THE PANCREATIC ISLET TISSUE (INTERALVEOLAR CELL ISLETS)

The pancreatic islet tissue, first described by Langerhans in 1869 consists of a large number of discrete groups of cells widely scattered throughout the pancreas (Fig. 17, 3) and accounting for between one and two per cent. of its total volume. The function of this tissue is apparently quite independent of that of the acini and in certain species of fish the islet tissue lies separately in the body. There are two main types of cells in the islets (Fig. 51, 15). The α-cells, about one-quarter of the total, are large cells with large spherical nuclei and granular cytoplasm; the smaller β-cells, lying mainly at the centre of the islets and accounting for 50 to 75 per cent. of the volume of each islet, contain large numbers of acidophil granules believed to be the hormone insulin itself. The islets of

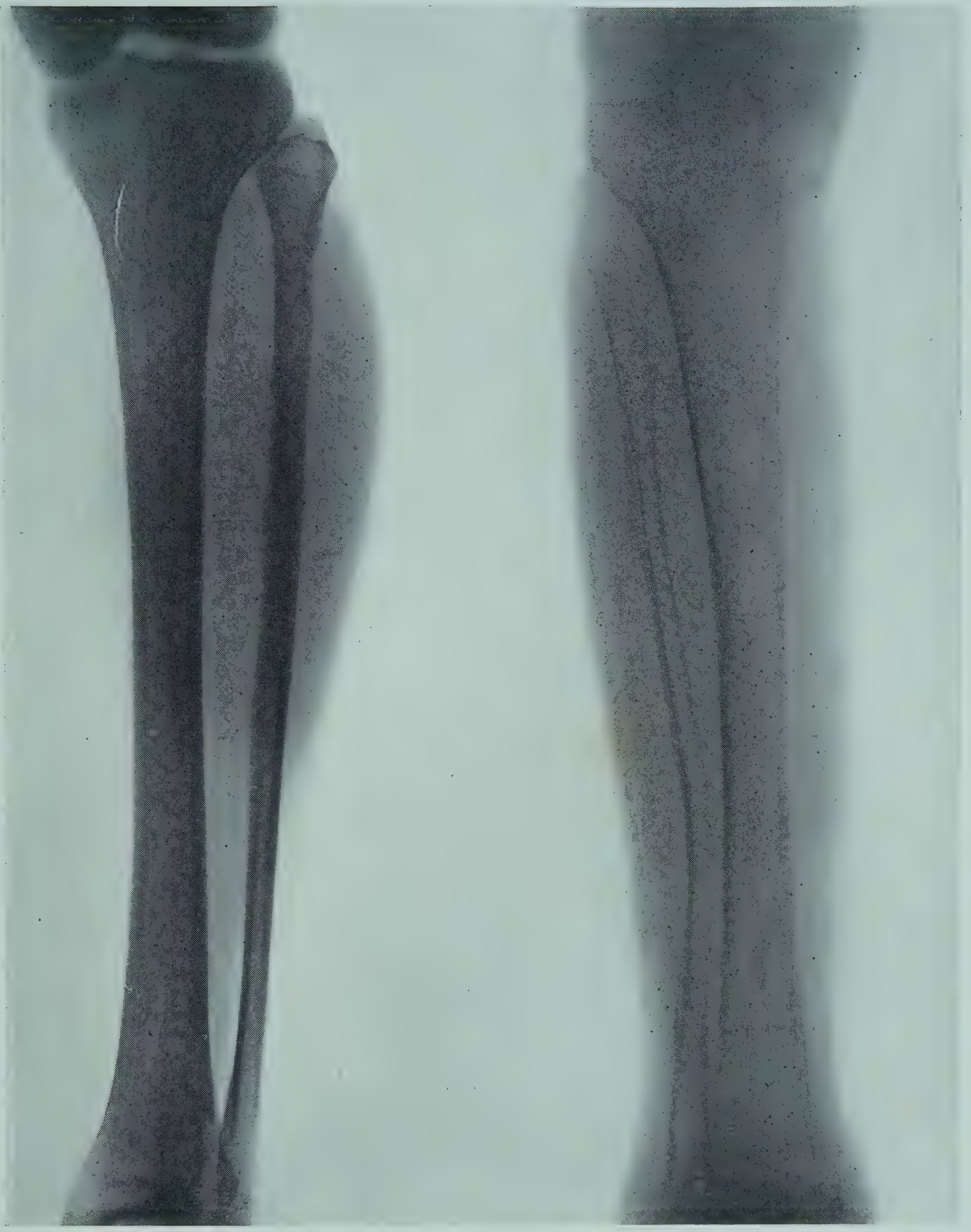
Langerhans contain relatively large amounts of zinc which increase during hunger and decrease after a carbohydrate meal; insulin may be in combination with zinc in the β-cells.

(a) **Insulin.**—The important role of the pancreas in carbohydrate metabolism was first demonstrated in 1889 by von Mering and Minkowski who showed that the removal of the pancreas from dogs was followed by a condition, similar to diabetes mellitus in man, in which large amounts of sugar were excreted in the urine. The idea that diabetes is a 'deficiency disorder' due to the absence of an unknown internal secretion from the pancreas was suggested by experiments made a few years later in which it was found that the condition was much relieved or completely cured by grafting pancreatic tissue into some other part of the body. In 1913 Allen reported that partially depancreatized dogs became permanently diabetic if they were given a diet rich in carbohydrate but not if they were starved or fed mainly on fat. The development of the diabetic state was accompanied by swelling of, and the appearance of vacuoles in, the β-cells of the islet tissue (hydropic degeneration) in the remaining portion of the pancreas. These experiments linked carbohydrate metabolism to the β-cells of the pancreatic islets and suggested that sustained hyperglycæmia can itself lead to damage to the islet tissue. It has since been shown that hyperglycæmia can injure the insulin-secreting cells to the extent that their ability to secrete insulin in adequate amounts is impaired. There is also evidence that injections of insulin can lead in certain circumstances to a fall in the insulin content of the pancreas, atrophy of the islet cells and permanent diabetes. The importance of these observations in relation to the treatment of diabetes is apparent.

Attempts to extract from the pancreas the anti-diabetic material were unsuccessful, until in 1921 Banting had the idea that the failures might be the result of the destruction, during the processes of extraction, of the internal secretion by the digestive enzymes of the external secretion. Along with Best in Macleod's laboratory in Toronto he produced degeneration of the acinar part of the gland by ligating the pancreatic duct and succeeded in extracting active material from the islet tissue (Fig. 51, 16).

The work of Banting and Best was extended by their colleague Collip, who in 1922 applied methods of protein fractionation to pancreatic extracts and produced crude preparations of the hormone insulin suitable for injection into animals and into diabetic patients. Their methods were soon applied on a commercial scale. A few years later (1926) crystalline insulin was first obtained, and to-day all insulin manufactured for clinical use is first purified to the stage of crystallization. The crystalline material usually obtained contains a small amount of zinc but active material can be obtained free from the metal.

Insulin is a protein with isoelectric point at pH 5·3 and a molecular weight of 6000. Polymerization to a molecular weight of 48,000 can, however, occur. The full structure of insulin as determined by Sanger is given on p. 50. It has been shown that its biological activity is a function of the whole molecule rather than of a specific grouping within it. Even slight alterations in molecular structure cause loss of potency. The proteolytic enzymes rapidly inactivate insulin and no satisfactory way has yet been found of protecting it from hydrolysis in the alimentary tract. Insulin must, therefore, be administered by injection.



A B

FIG. 51, 14

X-ray photographs of the right lower legs of two females, aged 25. A is from a normal subject, B from a patient with a parathyroid tumour. In B the fibula is deformed and both bones cast a poor shadow because of extensive reabsorption of calcium salts.

(By courtesy of W. T. Cooke.)

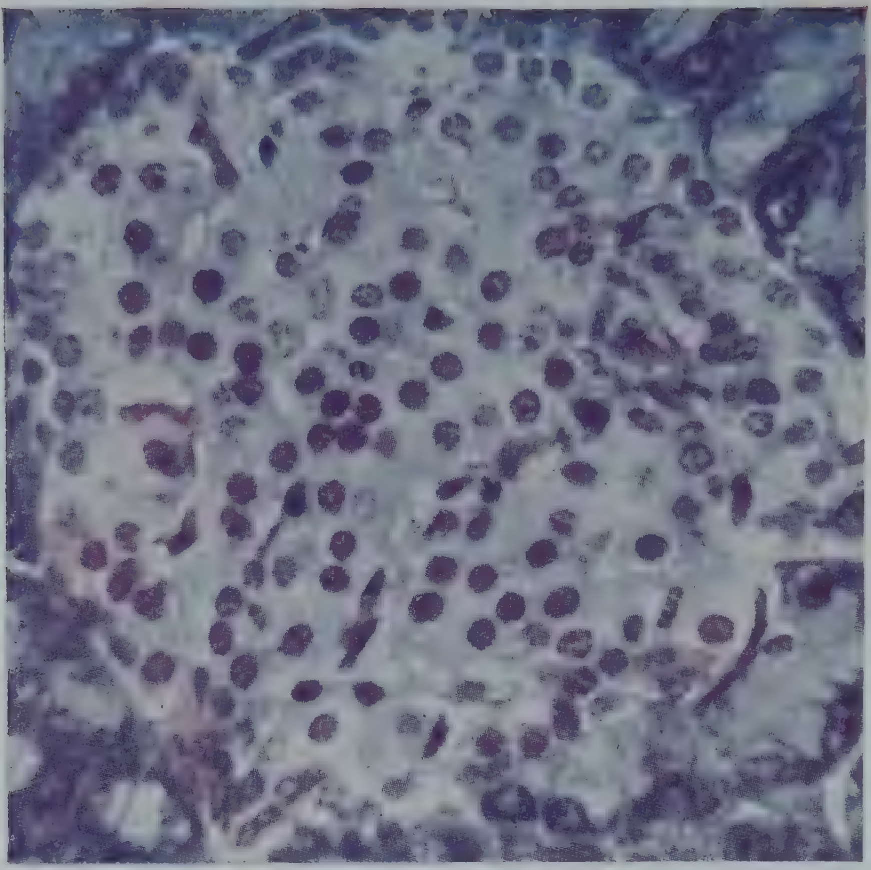


Fig. 51, 15a

From a control (healthy) rat showing relatively few α-cells (with red cytoplasm) mainly at the edge of the islet which contains a large number of β-cells. M. K. MacDonald & S. K. Bhattacharya (1956), *Quart. J. exp. Physiol.* **41,** 153–161. ×500.

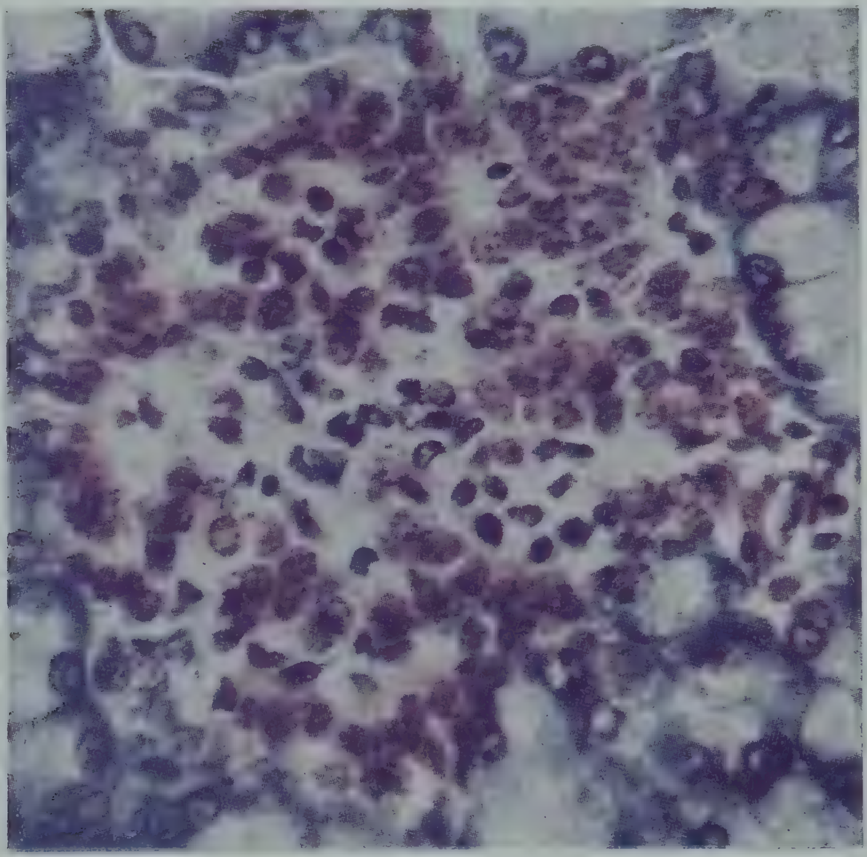


Fig. 51, 15b

From a rat made hyperglycæmic by repeated intravenous injections of dehydro-ascorbic acid. There are relatively few β-cells remaining towards the centre of the islet but numerous α-cells surrounding them. M. K. MacDonald & S. K. Bhattacharya (1956), *Quart. J. exp. Physiol.* **41,** 153–161. ×500.

Stain in both A and B: Gomori's chrome alum hæmatoxylin and phloxin.

Insulin is fairly soluble in tissue fluids and in plasma and rapidly diffuses away from the site of injection. It must therefore be given to diabetic patients several times a day, and even so the blood sugar level varies greatly throughout the twenty-four hours. This difficulty has been largely overcome by combining insulin, which is acidic, with a basic protein, protamine (p. 53) from fish sperm. This preparation is less soluble in body fluids than insulin itself and therefore exerts a prolonged action. The addition of zinc (protamine zinc insulin) further delays the action so that a single daily injection may serve to keep the blood sugar within satisfactory limits. Preparations of insulin containing zinc have also been devised in which the duration of action is related to the size and form of the particles in suspension (Table 51, 17).

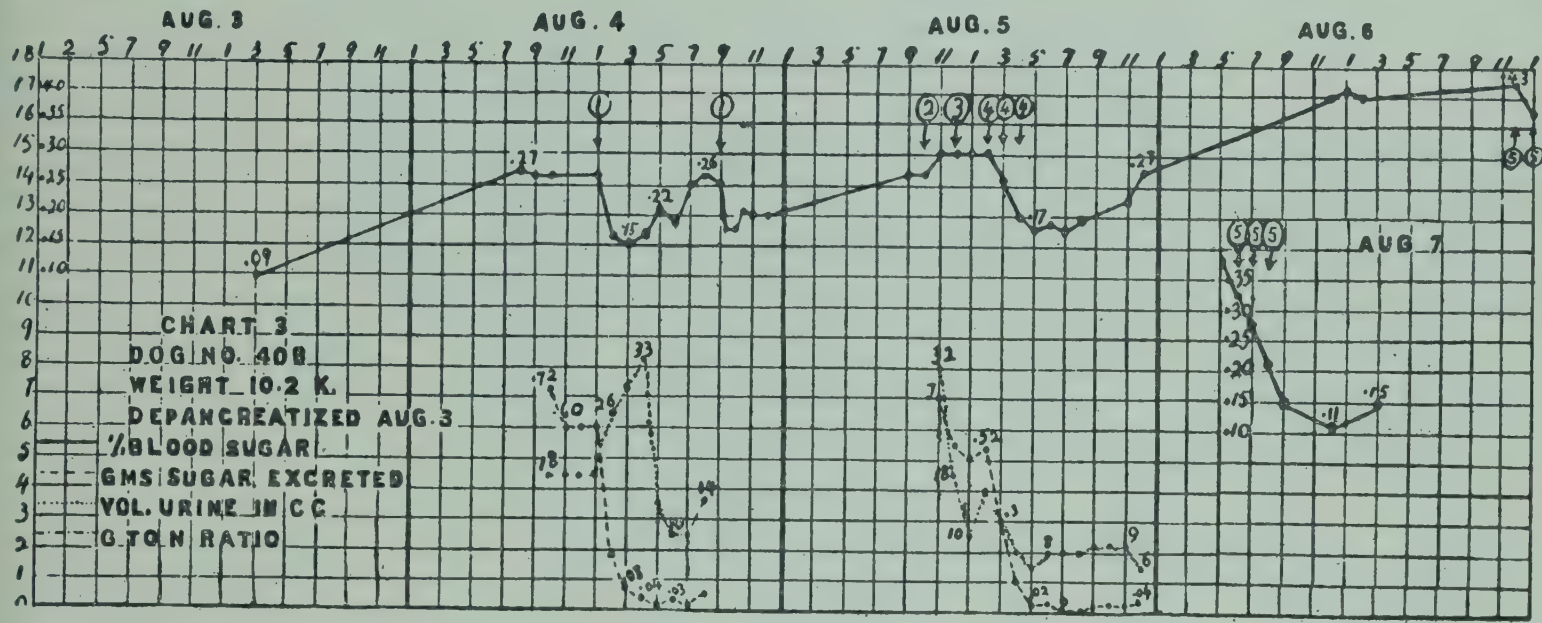


FIG. 51, 16

Original record of a classical experiment carried out by Banting and Best on a depancreatized dog. The normal blood sugar was 0·09 g. (90 mg.) per 100 ml. blood. Pancreatectomy was carried out on 3rd August and 18 hours later the blood sugar had risen to 0·27 g. per 100 ml. At ①, 5 ml. of a 4-day-old extract of degenerated pancreas were injected, causing a sharp fall in the blood sugar level. ②, ③, and ④ indicate injections of 5 ml. of extract of liver, extract of spleen, and extract of degenerated pancreas respectively. Only extract of pancreas produced a fall in the blood level. Further extracts of degenerated pancreas were injected at ⑤. The dog died on 12th August of general peritonitis (F. Banting & C. Best (1921), *J. lab. clin. Med.* **7**, 251).

With the preparation of increasingly active extracts of the pancreas it became important to have a standard to which the potency of such preparations could be referred. The present International Standard preparation of insulin set up in 1935 has been assigned an activity of 22 units per mg. but many of the preparations available now are more active than the standard. No chemical method of estimating the potency of insulin has yet been devised and biological methods must therefore still be used. Solutions of insulin are assayed by one of two methods: either by comparing the effects of standard and unknown in reducing the blood sugar level of rabbits, or by finding that dose of the test solution which produces hypoglycæmic convulsions in the same proportion of a group of mice as the standard solution.

Insulin is presumably destroyed in the body; it has never been detected in urine although its molecular size suggests that it might pass through the glomerular membrane. Insulin activity in the plasma can be assayed.

Perfusion of rat pancreas with blood containing a high concentration of glucose (140 to 570 mg./100 ml.) results in the appearance of insulin in the perfusate but, although this and other evidence suggests that hyperglycæmia stimulates the secretion of insulin, it is unlikely that this is the only way in which the output of insulin is regulated. Although nerve filaments from the vagus have been traced to the islet tissue and stimulation of the vagus leads to increased output of the hormone, the carbohydrate metabolism in animals in which the pancreas has been denervated or even transplanted is apparently normal. Injection of certain anterior pituitary extracts into rats has been followed by a hypertrophy of islet tissue and an increase in the insulin content of the pancreas but it seems probable that any influence by the anterior pituitary on insulin production is indirect.

TABLE 51, 17

Approximate Duration of Action of various preparations of Insulin

Preparation	Effect		
	Begin	Max.	End
Soluble insulin . .	30 minutes	2–4 hours	6 hours
Insulin zinc suspension (amorphous) . .	1 hour	8 hours	12 hours
Protamine zinc insulin .	6 hours	8–12 hours	26 hours
Insulin zinc suspension (crystalline) . .	4 hours	12–18 hours	30 hours

The actions of insulin have been described on pp. 301 to 304. To account for them are three theories of insulin action:

(1) Insulin potentiates the activity of hexokinase by counteracting the inhibitory influence of the adenohypophysis and adrenal cortex.
(2) Insulin is necessary for oxidative phosphorylation.
(3) Insulin promotes the transfer of glucose across cell membranes.

Each of these theories could explain the main action of insulin, namely its ability to increase the utilization of glucose. Until more knowledge is available it is impossible to decide between them; it may well be that the manifold effects of insulin cannot be ascribed to a single basic action.

Complete pancreatectomy in man, an operation occasionally performed for the eradication of malignant disease, is followed by diabetes mellitus which, however, is much milder than diabetes as usually met with in man. While the exact underlying mechanism of the latter is not yet known, it is already certain that deficiency of insulin is by no means the most important factor. However, the depancreatized dog and the untreated diabetic patient both exhibit hyperglycæmia and glycosuria, polyuria, ketonæmia and ketonuria and acidosis. These biochemical changes are accompanied by thirst (to replace the water eliminated with the glucose), hunger (because of the loss of sugar in the urine), fatigue and loss of weight (due to combustion of fat and protein in place of glucose), and finally increasing drowsiness, coma and death. Whatever may be the initiating cause of the diabetes, insulin given in sufficient amount and at frequent enough intervals can reverse these changes. Diabetes is dealt with in more detail on pp. 307 to 311.

Very occasionally an actively secreting tumour (adenoma) of the islet tissue occurs. Its excessive secretion of insulin (*hyperinsulinism*) produces the effect of hypoglycæmia, namely irritability and feelings of apprehension and nervousness, hunger, profuse perspiration, flushing alternating with pallor, and finally giddiness, convulsions and death. Since the brain utilizes carbohydrate almost exclusively, the lowered level of blood sugar is accompanied by reduction of cerebral metabolism, alteration of electroencephalogram (p. 907) and unconsciousness.

(*b*) **Glucagon.**—The Hyperglycæmic-Glycogenolytic Factor. — This factor, which is also a protein, is produced by the α-cells of the pancreatic islets. It raises the blood sugar level and is present in most commercial insulin preparations although its action is overshadowed by the hypoglycæmic effect of the insulin. It has been discussed on page 304.

THE SUPRARENAL GLANDS

The two suprarenal or adrenal glands are situated one on each side of the vertebral column, closely applied to the upper poles of the kidneys with which they are enclosed in the renal fascia. These glands contain two different kinds of tissue, *cortex* and *medulla* (Fig. 51, 18), which have little physiological relationship and indeed have quite different embryological origins. Complete removal of both glands is fatal within a few days but all the medullary tissue can be extirpated with little apparent effect; the adrenal cortex is therefore essential to life, while the medulla is not.

Anatomically the glands are remarkable for their extreme vascularity and for the fact that each is supplied with blood by three relatively large arteries, the branches of which enter the gland at its periphery. Having passed through the capillaries of the cortex, the blood enters sinusoids in the medulla and leaves the gland by a single large vein emerging from the hilum.

The suprarenal glands are supplied by a large number of preganglionic nerve fibres which run without synaptic interruption in the splanchnic nerves to end in close relation to the cells of the medulla which are therefore homologous with the cells of sympathetic ganglia. The suprarenal cortex receives no nerve connexions.

THE SUPRARENAL OR ADRENAL CORTEX

The suprarenal cortex is of mesodermal origin, being derived from the intermediate cell mass from which develop the primitive kidney and the gonads. In man it is not uncommon to find small masses of tissue identical with the suprarenal cortex ('cortical bodies') in the neighbourhood of the gland or even in other situations. Microscopic examination of the cortical tissue shows it to be made up of three more or less distinct zones (Fig. 51, 18) but it is uncertain whether there are any functional differences between them. The cortical cells, especially in the zona fasciculata, are rich in lipids which appear as granules in the cytoplasm; their remarkably high content of ascorbic acid, 200 mg. per 100 g., makes them stain black with silver nitrate. Administration of adrenocorticotrophic hormone (ACTH, p. 1004) reduces the amount of lipid throughout the cortex and induces hypertrophy of the zona fasciculata and zona reticularis; it accelerates the rate of secretion of cortical hormones.

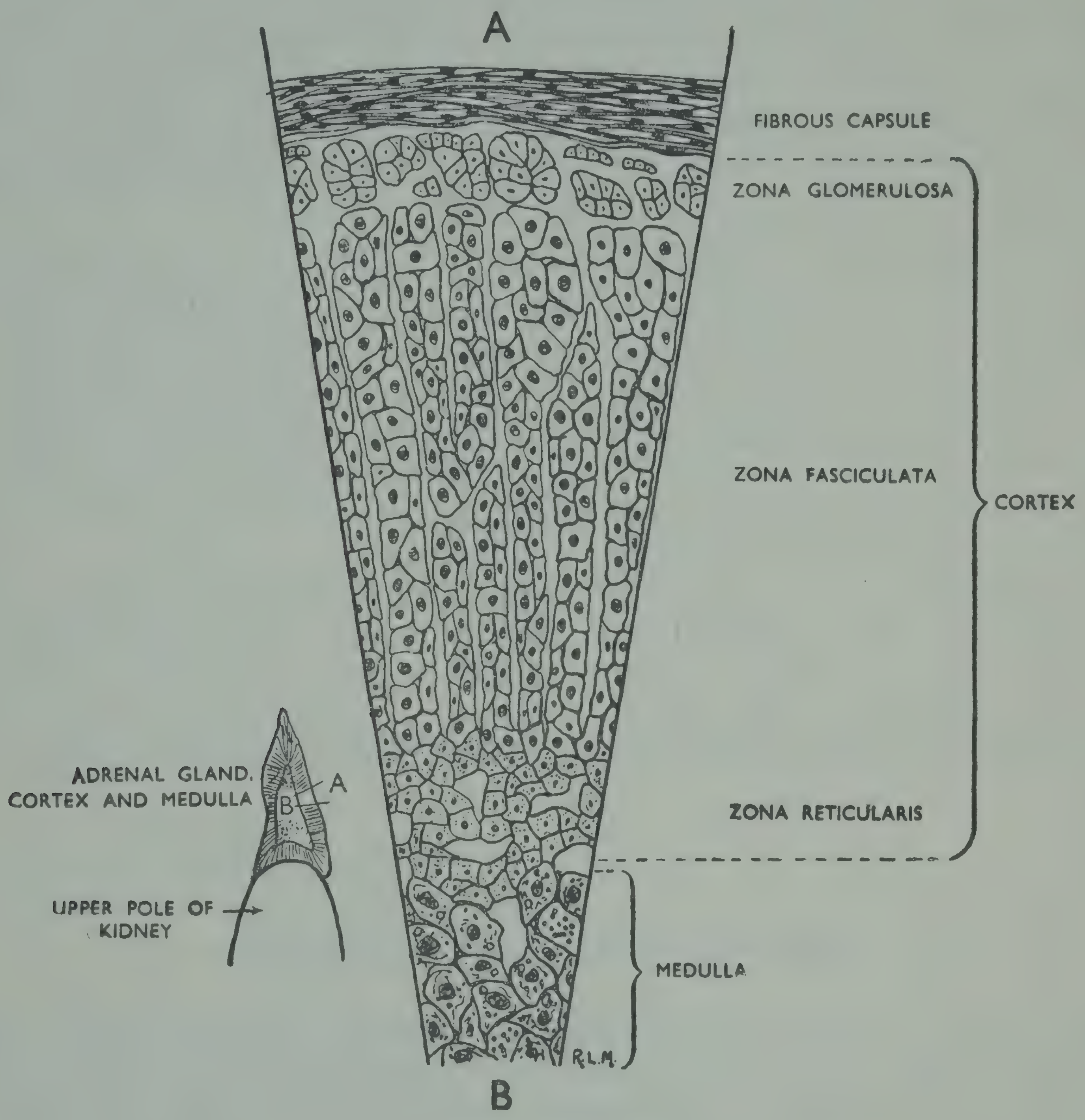
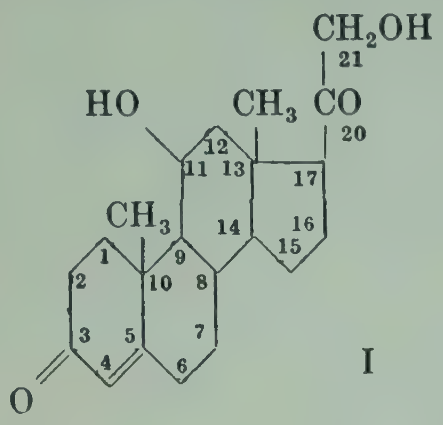

A
FIBROUS CAPSULE
ZONA GLOMERULOSA
ZONA FASCICULATA
CORTEX
ZONA RETICULARIS
MEDULLA
ADRENAL GLAND. CORTEX AND MEDULLA
A
B
UPPER POLE OF KIDNEY
R.L.M.
B

Fig. 51, 18

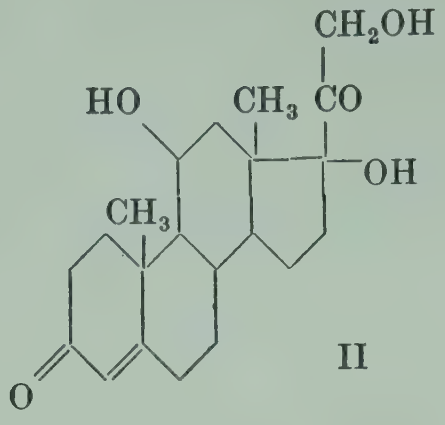
Diagram of adrenal gland showing differentiation into medulla and cortex, the latter itself being divided into three zones not clearly demarcated from one another.

Within 48 hours of extirpation of both suprarenal glands in a dog, the animal becomes severely ill and, if untreated, it succumbs in 5 to 10 days. Soon after the operation there is loss of appetite and rapid loss of weight with profound weakness and prostration. The plasma volume is diminished and the animal soon becomes severely dehydrated; in the plasma the concentration of sodium and glucose falls, while there is progressive increase in the urea, non-protein nitrogen, potassium and magnesium. The striking increase in the serum magnesium concentration may be responsible for the muscular weakness, since magnesium ions inhibit the breakdown of adenosinetriphosphate. Vomiting and diarrhœa, with


CH$_2$OH
21
HO
CH$_3$
CO
20
12
11
13
17
CH$_3$
14
16
15
1
9
2
10
8
3
5
7
4
6
O

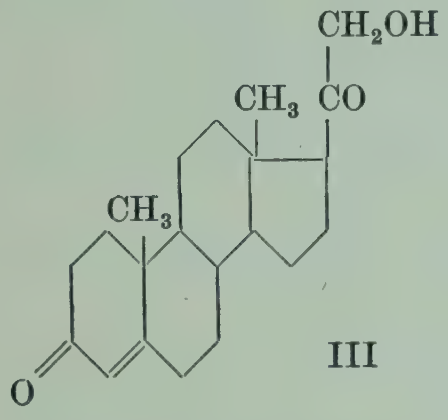
I

Corticosterone
(compound B)


CH$_2$OH
HO
CH$_3$
CO
OH
CH$_3$
O

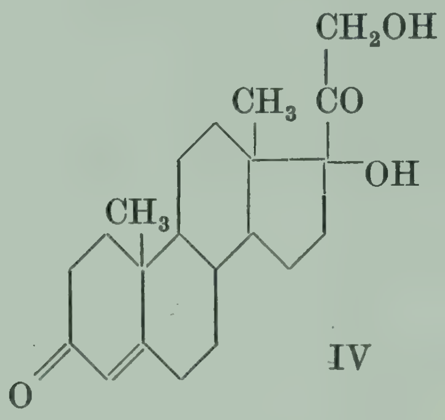
II

17-Hydroxycorticosterone
(cortisol, compound F)


CH$_2$OH
CH$_3$
CO
CH$_3$
O

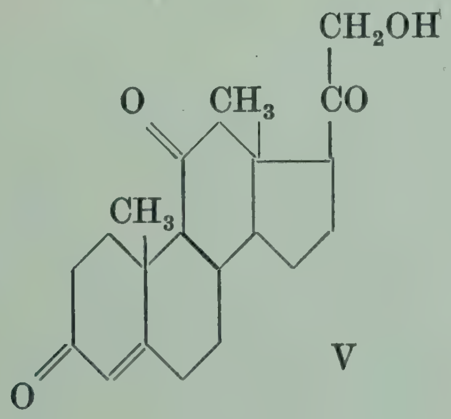
III

11-Deoxycorticosterone
(cortexone)


CH$_2$OH
CH$_3$
CO
OH
CH$_3$
O

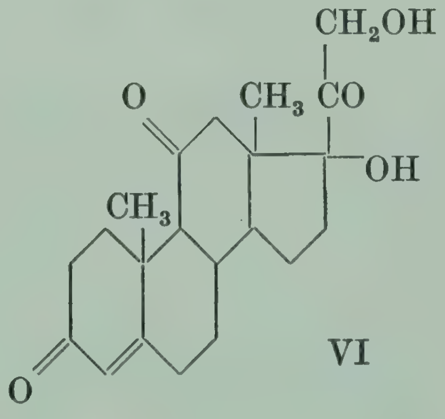
IV

11-Deoxy-17-hydroxycorticosterone


CH$_2$OH
O
CH$_3$
CO
CH$_3$
O

V

11-Dehydrocorticosterone
(compound A)


CH$_2$OH
O
CH$_3$
CO
OH
CH$_3$
O

VI

11-Dehydro-17-hydroxycorticosterone
(cortisone, compound E)

Fig. 51, 19

Six of the steroid hormones of the adrenal cortex. They are all named after the parent substance corticosterone which has a hydroxyl group on position 11. V has a keto group on this position whereas III has no oxygen atom at this site and is therefore known as a deoxy compound. II, IV and VI correspond to I, III and V respectively, but with the addition of a hydroxyl group on position 17.

a fall in body temperature and blood pressure and a decrease in metabolic rate, lead finally to coma and death.

It is not at present clear how all these effects are related but Loeb emphasizes the primary importance of increased elimination of water, and more especially of sodium, in the urine, the fall in serum sodium concentration and the extreme dehydration. This cannot, however, apply to all animals since bilateral adrenalectomy in the opossum, though fatal, does not result in salt depletion. Adrenalectomized dogs can be maintained in good health indefinitely by the injection of cortical extract, particularly if a large amount of salt is included in the food, but withdrawal effects rapidly occur when injections are withheld (Fig. 51, 21). Cortical extracts are effective when given by mouth but very much larger doses are required.

The first potent extract of suprarenal cortex was made by Swingle and Pfiffner in 1930 and since then at least thirty different steroids have been isolated from cortical tissue. It is difficult to imagine that so large a number of hormones can be produced by a single gland and it may be, as Hartman has suggested, that the crystalline compounds so far isolated may be the more stable derivatives of quite a small number of less stable hormones. For purposes of description the hormones can be divided into four groups.

(1) Steroids of the group typified by deoxycorticosterone, the so-called *mineralocorticoids* (III, Fig. 51, 19), have no oxygen atom at position 11. Their action is predominantly the retention of sodium and water as the result apparently of a specific effect upon the proximal tubules of the kidney. Following their administration the plasma concentration of sodium increases and that of potassium falls. At the same time the concentration of potassium in the tissue cells is elevated and that of sodium is diminished.

(2) A second group of cortical hormones (the *glucocorticoids*) is exemplified by cortisone (VI, Fig. 51, 19). This group possesses an oxygen atom at position 11 and has its main action on carbohydrate metabolism (p. 305); it has little direct action on salt and water metabolism. The glucocorticoids interfere with the formation of antibodies and inhibit the normal response of body tissue to inflammation. Some 60 per cent. of the total production of steroids by the adrenal cortex can be accounted for by corticosterone and cortisol.

(3) Crude extracts of adrenal cortex after removal of steroids belonging to the two groups described above contain a substance which is capable of restoring renal function and of maintaining life after adrenalectomy. The active substance is *aldosterone* (*electrocortin*) (Fig. 51, 20). Biological tests show it to be some 50 to 100 times more active than deoxycorticosterone in promoting the retention of sodium.

(4) The steroids of the fourth group that have been isolated from adrenal cortex are related to the sex hormones. Both œstrogenic and androgenic compounds, as well as progesterone, are elaborated by the cortex. The chief member of this group is adrenosterone, the structure of which is closely similar to that of testosterone, but a variety of androgenic substances has been isolated from the urine of patients with tumours of the suprarenal cortex.

It has recently been shown by Pincus that the adrenal gland can synthesize steroids from cholesterol in the following order: cholesterol ⟶ progesterone ⟶ cortexone (III) ⟶ corticosterone ⟶ cortisol (II). ACTH is required for the first step but probably not for the later stages which involve hydroxylation at positions 11, 17 and 21.

The possible role of the adrenal cortex in fat metabolism was mentioned on p. 272. Adrenalectomized animals are unable to utilize their fat deposits as quickly as normal animals and adrenal insufficiency decreases the ability to absorb lipids from the alimentary tract, especially the long chain fatty acids.

One of the striking features of adrenalectomized animals is their inability to react rapidly to alterations in their environment. Adrenalectomized dogs and rats, for example, are unable to maintain their body temperature when placed in a cold room and patients with adrenocortical insufficiency are unduly susceptible to infections and anoxia and to many potentially toxic agents such as anæsthetics and morphine. Selye has recently studied the process of adaptation in animals continuously exposed to stress, such as cold. He has described the 'general adaptation syndrome' which includes "the sum of all those non-specific systemic reactions of the body which ensue upon long-continued exposure to stress." It includes a number of morphological and functional changes in the endocrine glands, but especially in the adrenal cortex. Stress, the injection of infective or toxic materials, exercise or hæmorrhage, stimulates the adrenal cortex provided that the hypophysis is present, presumably through the agency of ACTH. The output of cortical hormone is thereby increased and the cortex stimulated to hypertrophy. The means by which the adenohypophysis is stimulated by these agencies is doubtful; adrenaline has this effect, but stress produces cortical effects even when the adrenals are denervated. It may well be that the stimulation is effected through the hypothalamus.

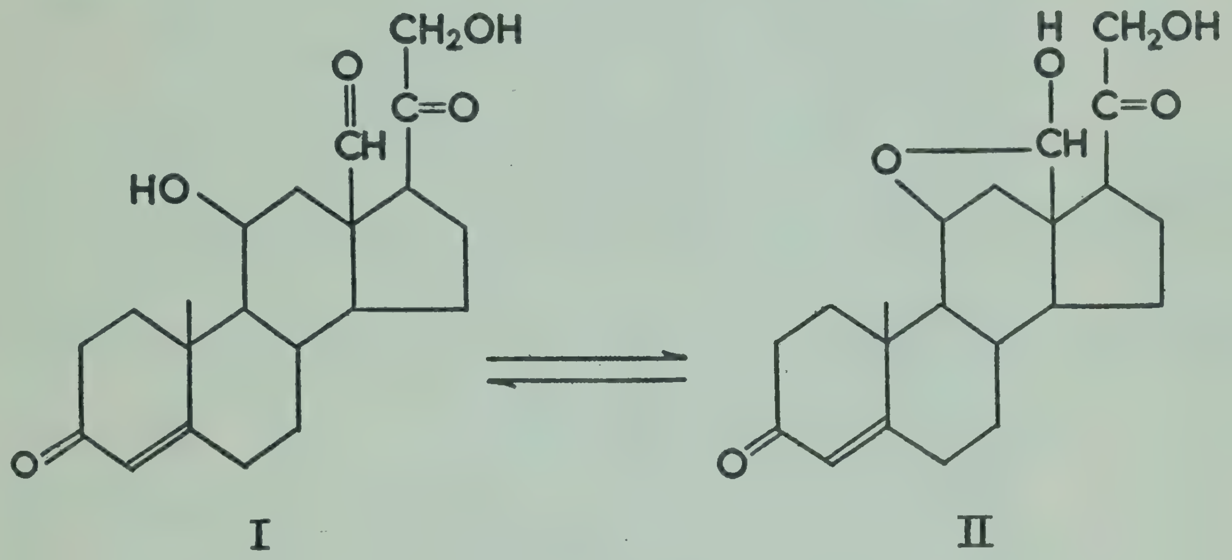


FIG. 51, 20

Aldosterone.
$\triangle^4$-pregnene-11β, 21-diol-3, 20-dione-18-al [Formula I].
In solution it appears to be in equilibrium with the 11, 18-hemiacetal [Formula II].

Destruction of the suprarenal glands by disease in man produces *Addison's disease*, which is the chronic counterpart of the acute disorder produced in animals by removal of both suprarenals. The chief symptoms as described by Addison in 1855 are "anæmia, general languor and debility, remarkable feebleness of the heart's action, irritability of the stomach and a peculiar change in the colour of the skin." These symptoms are accompanied by a low blood pressure, loss of weight and the biochemical features already described as following adrenalectomy in animals. The brown colour of the skin is due to the deposition of granules of melanin in the basal layers of the epidermis. It has been assumed that the excessive production of melanin, presumably from tyrosine, is a sequel to the destruction of the suprarenal medulla in this disease and the consequent impaired adrenaline production. Patients with Addison's disease can often be kept in good health for a long time if they take sodium chloride by mouth, but many also require injections of cortical extract or of deoxycorticosterone acetate (DCA). By implantation in tablet form of 300 mg. of DCA, from which approximately 1 mg. of DCA is absorbed per day, a patient may be kept comfortable for six or eight months and with repeated implantations he may live for many years. The administration of cortisone daily by mouth elevates the fasting blood sugar concentration

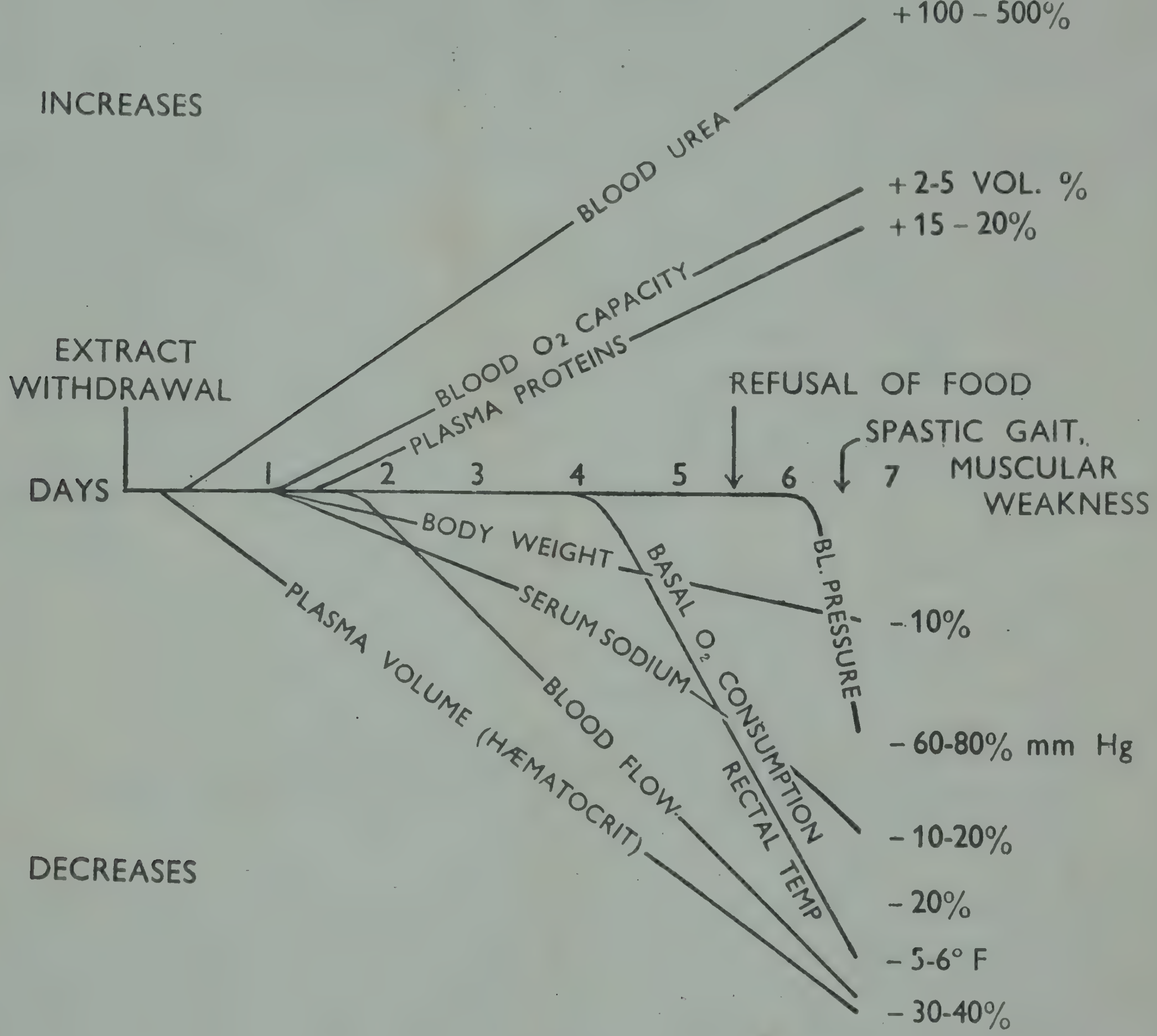


FIG. 51, 21

Chart showing graphically the development of acute cortical insufficiency following withdrawal of treatment in a group of adrenalectomized animals previously maintained in good health by administration of cortical extract (G. A. Harrop, A. Weinstein, L. J. Saffer, & J. H. Throsher (1933), *J. exp. Med.* **58,** 1).

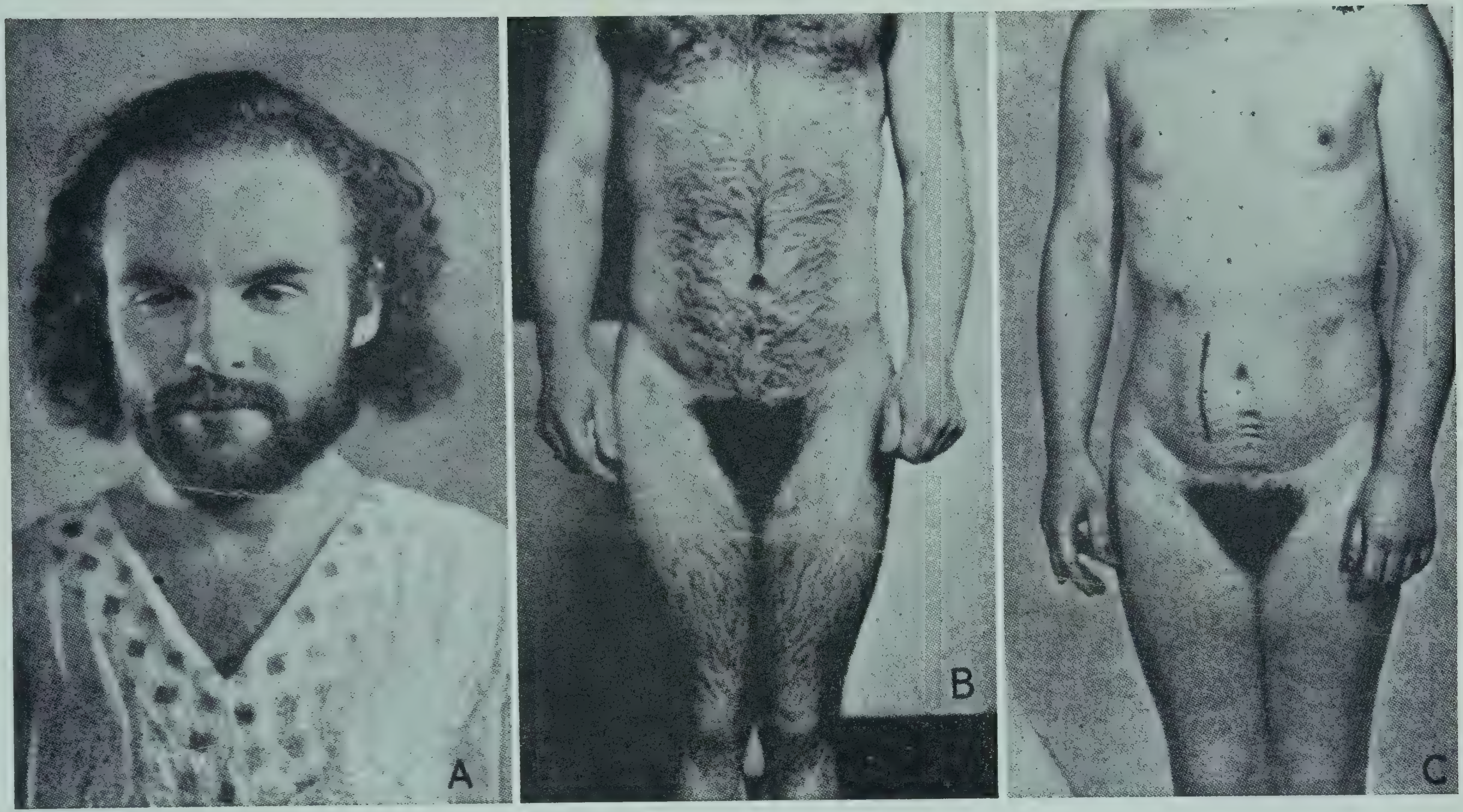


FIG. 51, 22

Adrenocortical virilism in a female patient, aged 17, with an adrenal cortical tumour. A and B were taken before operation; C, two years and ten months after removal of the tumour (F. D. W. Lukens & H. D. Palmer (1940), *Endocrinology*, **26,** 941).

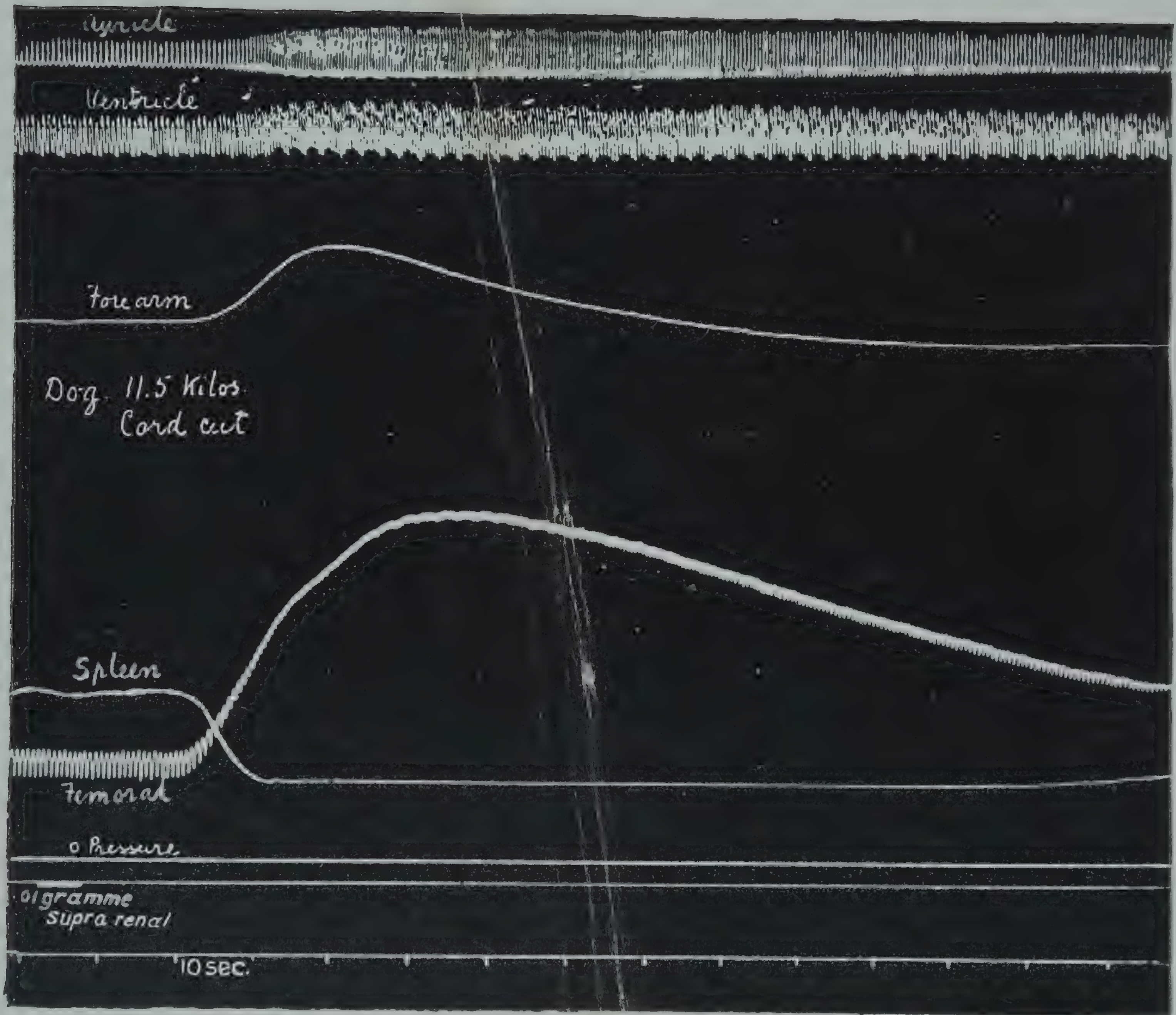


Fig. 51, 23

Reproduction of an original figure of G. Oliver and E. A. Schäfer (1895), *J. Physiol.* **18,** 246, to show the effect of an extract of the suprarenal in increasing the rate of the heart, elevating the arterial pressure, causing decrease of splenic volume and increase in the volume of the forelimb. Time marks every 10 sec.

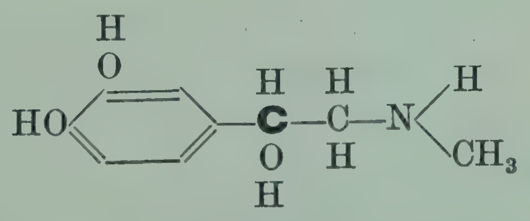
and improves the well-being of patients with Addison's disease. Procedures for the recognition of Addison's disease have been based on the findings that such patients do not have a normal diuresis after drinking a large quantity of water (a feature that is usually restored by cortisone) and secondly that they tend to excrete excessive amounts of sodium chloride but to retain urea.

Hyperfunction of the adrenal cortex in children, the result of either cortical overactivity or of an actual tumour of the cortex, results in precocious sexual development, premature puberty, hirsutism and muscular hypertrophy (the 'infant Hercules'). Hyperfunction of the adrenal cortex occurring in the adult may produce the *adrenogenital syndrome* of which one of the most striking features in the female is virilism (Fig. 51, 22). A similar syndrome is produced by overactivity of the anterior lobe of the pituitary. Assay of androgenic substances (17-ketosteroids, 17-oxosteroids) in the urine is an important diagnostic test in these conditions in which, of course, the urinary output tends to be much increased.

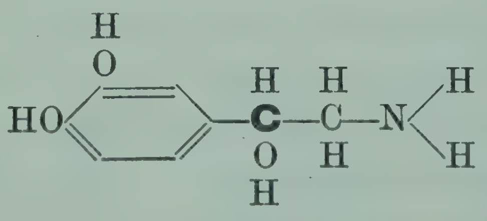
THE SUPRARENAL OR ADRENAL MEDULLA

The suprarenal medulla, which has a common origin with the ganglion cells of the sympathetic nervous system from the primitive neuro-ectoderm, consists of masses of polyhedral cells separated by large blood sinuses. The cells contain granules stained blue by ferric chloride or brown with salts of chromic acid (*chromaffin tissue*). The suprarenal medulla is richly supplied by a plexus of sympathetic nerve filaments derived mainly from the greater splanchnic nerve. The innervation is remarkable in that the plexus consists entirely of myelinated preganglionic fibres, the cells of the medulla representing the sympathetic ganglion cells and their postganglionic fibres.

The active principles of the medulla are easily extracted and when injected into animals they produce a sharp rise of blood pressure. It is not surprising, therefore, that this was one of the first hormonal effects to be discovered. In 1894 Schäfer and Oliver injected a watery extract of the suprarenal into a cat and obtained an immediate large rise of blood pressure which was, however, transient (Fig. 51, 23). It was later found that this property belonged to the medulla only and in 1900 one active principle was isolated and named *adrenaline* (or epinephrin). Recently it has been found that normal suprarenal tissue from man and many other animals contains *noradrenaline* (arterenol) as well as adrenaline and they are often referred to together as *catechol amines.* The ratio of adrenaline to noradrenaline in fresh human adrenals is 4 : 1.



adrenaline



noradrenaline

Adrenaline has an asymmetric carbon atom (**C**) and therefore exists in two optically active forms. That prepared from the gland is L-adrenaline which is some fifteen times more active in raising the blood pressure of the spinal cat than is the D-isomer. Synthetic adrenaline is of course racemic but the mixture is resolved after preparation and the L-form is supplied for medical use, usually as a 1 : 1000 solution of the hydrochloride or bitartrate. The commercial solution, if prepared from ox adrenals, may contain up to 20 per cent. of noradrenaline. Adrenaline in solution is very readily oxidized and inactivated but blood and tissue fluids contain reducing agents, such as glutathione and ascorbic acid, which can protect it from rapid oxidation. However, adrenaline is rapidly destroyed in the body by the cytochrome oxidase system with the formation of adrenochrome or by the amine oxidases which convert it to methylamine. Calcium enhances and potassium antagonizes some actions of adrenaline.

Although the steps by which adrenaline is synthesized by the chromaffin tissue are not yet clear the close chemical relationship between adrenaline and tyrosine suggest the latter as the most probable precursor. Experiments on rats with the isotope ^{13}C have shown that phenylalanine, which can be converted to tyrosine (p. 348), is the precursor of both noradrenaline and adrenaline in the suprarenal glands.

The medullary secretion is under the control of the splanchnic nerves which are cholinergic, that is, they act by producing acetylcholine at their endings in the medulla. However, recent experiments on the denervated perfused gland show that a period of anoxia leads to a discharge of adrenaline and it may be presumed that anoxia has a direct effect on the cells of the suprarenal medulla. The function of the adrenal medulla is apparently unaffected by removal of the anterior lobe of the pituitary.

Neither complete removal of all medullary tissue nor denervation of the suprarenals, which reduces the output of adrenaline and noradrenaline, handicaps the animal seriously if it is not exposed to stress. In other words the suprarenal medulla, though important in an emergency and under adverse environmental conditions, is not essential to life.

The humoral transmission of the effect of the nerve impulse from most sympathetic fibres as the result of the liberation of an adrenaline-like substance at the sympathetic nerve-endings has been dealt with in Chapter 43. The injection of adrenaline or noradrenaline is followed by effects all over the body; these are dealt with in more detail in Chapter 43 and summarized in Table 43, 5.

Since pure adrenaline and noradrenaline are readily available it is unnecessary to have a biological method for their standardization. In physiological research, however, it may be necessary to determine the amounts of noradrenaline and adrenaline in an extract or perfusate. This can be done by using organs which respond differently to the two substances. For example, the cat's blood pressure is more easily elevated by noradrenaline, while adrenaline produces a greater relaxation of the rat's uterus. It should be emphasized here that the vascular system of man is more sensitive to both adrenaline and noradrenaline than that of other species and this should be borne in mind whenever an attempt is made to apply to man conclusions drawn from animal experiments.

Our knowledge of the physiology of medullary secretion has been greatly increased recently as it is now possible to estimate the amounts of the catechol amines in the suprarenal vein of the cat. The venous effluent usually contains more noradrenaline than adrenaline. When the carotid arteries are clamped there is reflex vasoconstriction and the adrenal effluent contains mostly noradrenaline. If, however, a painful stimulus is applied, for example by stimulation of the sciatic nerve, more adrenaline is secreted. In the hypoglycæmia produced by insulin, adrenaline is secreted but in the hyperglycæmia produced by intravenous injection of glucose there is a reduction of the adrenaline output with no change in the noradrenaline output. Denervation of the cat's adrenal reduces the noradrenaline output to one-third. Folkow and von Euler have recently found that stimulation of one particular area of the hypothalamus causes mainly adrenaline to appear in the effluent whereas stimulation of another area of the hypothalamus releases mainly noradrenaline. All these findings suggest that adrenaline and noradrenaline are released separately and histochemical tests have produced evidence to support the idea that there may be two sets of cells in the adrenal medulla, one responsible for liberating adrenaline and the other for liberating noradrenaline.

TABLE 51, 24

Condition	Urinary excretion in mμg./min	
	Adrenaline	Noradrenaline
Rest	2–4	8–15
Mild exercise	2 times	2 times
Strenuous exercise . . .	10 times	10 times
Mental stress	Increased	No effect
Insulin hypoglycæmia . .	10 times	No effect
Tilt, recumbent to head up .	Little change	Increased
After adrenalectomy . .	0	8–15
Chromaffin cell tumours . .	Slightly increased	100 times or more

Urinary excretion of catechol amines in man. (From U. S. von Euler (1955), Noradrenaline in hypotensive states and shock, *Lancet*, ii, 151–154.)

The excretion of the catechol amines (Table 51, 24) provides further information about the activity of the adrenal medulla since 2 to 3 per cent. of injected catechol amines are excreted in the urine. The absence of adrenaline after adrenalectomy suggests that the urinary adrenaline comes from the medulla; the urinary noradrenaline output continues and must, therefore, come from extramedullary sources such as the adrenergic nerves. In strenuous exercise the increase in output is correlated with fatigue rather than with the amount of work.

On the basis of these findings von Euler makes a contrast between the two hormones as follows. Adrenaline, by producing vasoconstriction in the skin and vasodilatation in the muscles, moves blood from the skin to the muscles; it stimulates metabolism and mobilizes glycogen as glucose. These preparations are valuable for "fight or flight"—that is the emergency function as suggested by Cannon. Noradrenaline on the other hand produces a general vasoconstriction (except in the coronary arteries) and stimulates the heart but its effects on metabolism are inconspicuous. Noradrenaline may be regarded as a pressor hormone required for maintenance of blood pressure.

The methods of analysis are scarcely sensitive enough yet to measure the concentration of the catechol amines in peripheral blood—it must be less than 1 μg. per litre. About 30 μg. of L-noradrenaline and about 10 μg. of L-adrenaline are excreted by a healthy man each day.

Hypofunction of the adrenal medulla has not been recognized as producing any clear-cut clinical syndrome. Actively secreting chromaffin cell tumours, *phæochromocytomata* however, are known to produce large quantities of noradrenaline with a small amount of adrenaline, and their discharge at intervals into the blood may produce temporary elevations of arterial pressure. Up to 2 mg. of these substances may then appear in the urine per day.

THE THYMUS AND PINEAL BODY

The Thymus is an irregular two-lobed organ lying behind the sternum. Closely resembling a lymph node microscopically, it consists of numerous nodules or follicles 1 to 2 mm in diameter containing large numbers of lymphocytes with a few polymorphs. In the centre of the follicle the lymphocytes are less densely packed and occasional *concentric corpuscles of Hassall* are seen. The function of these structures is unknown.

The thymus gland is relatively large in the child and reaches its maximum size, about 35 g., at puberty, after which it regresses and is replaced by fatty tissue. For a long time it was accepted that enlargement of the thymus was associated with *status lymphaticus*, a condition in which sudden death may occur in apparently healthy children after very minor injuries. While it is doubtful if there is such a condition, and it is certainly not to be connected with the thymus, the origin of the mistake is interesting. At post-mortem examinations of children who have died as a result of an illness which has lasted for some days or weeks the thymus is always small, while in children who die suddenly, as after an accident, the thymus appears to be enlarged. It was not recognized at first that this apparently large thymus was in fact one of normal size. When there was no apparent explanation of the sudden death it was but natural to suppose that the large thymus was in some way implicated.

Rowntree described some remarkable experiments in 1935 in which long-continued injections of thymus extracts were given to successive generations of rats, with the result that the later generations reached maturity unusually rapidly. Since these results have not been confirmed the relation of the thymus to growth is still undecided.

Enlargement of the thymus gland has been found in cases of *myasthenia gravis*, in which all the skeletal muscles are weak because of a derangement of neuromuscular transmission. The surgical removal of the thymus has been frequently performed on these patients with apparent benefit but the exact connexion between myasthenia and the thymus remains obscure.

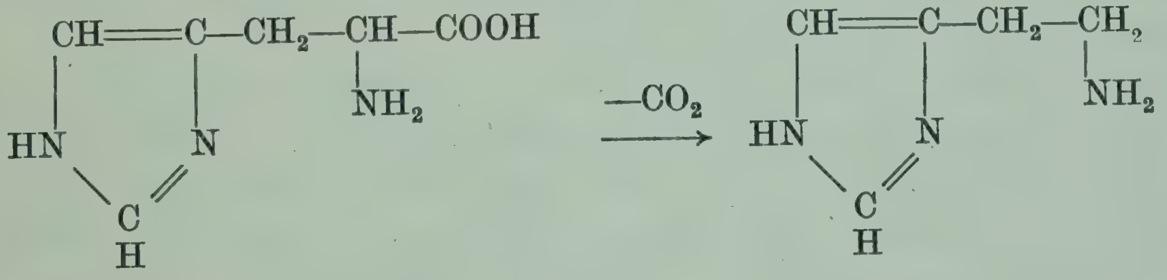
The Pineal body, often visible in X-rays of the skull owing to its tendency to calcify, is a small structure about 8 mm long and 4 mm wide lying on the dorsal aspect of the superior quadrigeminal bodies. It contains a cavity which communicates with the third ventricle. Although an œstrogenic substance has been extracted from the pineal body its function is quite unknown, conflicting results having been described following its removal in animals. Symptoms produced by tumours of the pineal body in man may well be due to pressure on adjacent structures.

LOCAL HORMONES

Great interest is being shown at the present time by physiologists and pharmacologists in substances which are released locally and produce local effects, such as vasodilatation or vasoconstriction, in the surrounding tissue. The so-called local hormones include histamine, 5-hydroxytryptamine, adrenaline, noradrenaline and acetylcholine. Some of these substances are, of course, also concerned in the humoral transmission of nerve impulses.

Acetylcholine (ACh).—Isolated rabbit atria can continue to beat *in vitro* for about 24 hours. During the early part of that time the addition of acetylcholine to the bath results in a depression of contraction. However, if acetylcholine is added to the bath when contractions have ceased, the atrial strips begin to beat again. This phenomenon is explained thus. The fresh atrium contains an enzyme system which synthesizes ACh; if ACh is added to the bath the synthesis is depressed and the beat weakens. In atria which have ceased to beat synthesis of ACh is greatly depressed but is revived by ACh and the muscle strip begins to beat again. The stimulating effects of ACh on the small intestine cannot, however, be explained on the basis of its effect on an enzyme system.

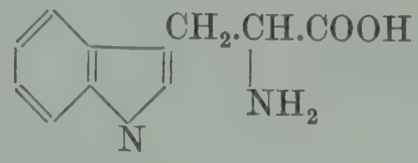
Histamine.—Large amounts of histamine are present in the lumen of the intestine due apparently to decarboxylation of histidine by bacteria.



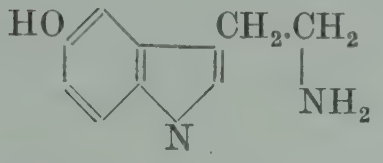
histidine histamine

It is present in many tissues in a combined inactive form. There is histamine in the skin, in skeletal muscle, in nerve fibres, in the wall of the intestine, and in the liver and lung. The submucosa and muscular coats of the small intestine contain relatively large amounts of histamine but it is difficult to account for this by absorption as the intestinal mucosa contains histaminase. In several animals there is a good correlation between the histamine content of a tissue and the number of mast cells (basophil leucocytes) it contains but these cells are not likely to be the only source of histamine. The mast cells also contain heparin which is released along with histamine when the animal is injected with a histamine liberator. The effects of histamine on the skin (triple response p. 520 and itching p. 658) and on gastric secretion (p. 247) have already been described. Urticaria, or nettle-rash, may be presumed to be due to a release of histamine. Gastrin (p. 242) may be a histamine liberator. But beyond these scraps of information we know very little about the physiological role of this substance.

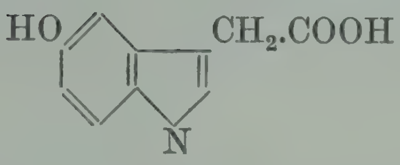
Hydroxytryptamine (Serotonin).—The serum obtained after blood clots has vasoconstrictor properties now attributed to 5-hydroxytryptamine (5 HT) which can be thought of as being derived from tryptophan by decarboxylation and addition of a hydroxy group. The 5 HT is released from the blood platelets when they disintegrate during clotting and it presumably aids the process of hæmostasis. Human blood contains 0·1 to 8·0 μg. per ml. of 5 HT, nearly all this being in the platelets. In animals clots injected intravenously produce pulmonary embolism with blockage of the pulmonary arterial bed and a marked fall of blood pressure, an effect which can be imitated by injection of 5 HT. The intestinal mucosa contains about 6 μg. per g. of 5 HT; it is present in the argentaffin cells. The brain contains about 0·2 μg. per g. Stimulation of some peripheral nerves, including the vagus, leads to release of 5 HT from the brain. Human urine normally contains 5 to 10 mg. of 5-hydroxyindoleacetic acid (5 HIAA), a derivative of 5 HT, per 24 hours. The intravenous injection of 5 HT in man causes a sharp rise in blood pressure by a direct action on the smooth muscle of the blood vessels.



tryptophan



5-hydroxytryptamine (serotonin), 5HT



5-hydroxyindoleacetic acid, 5HIAA

A large number of antimetabolites of 5 HT are known, the most important being lysergic acid diethyl amide (LSD). If 30 μg. of LSD are taken by mouth mental disturbances including hallucinations occur and it is tempting to speculate that these are brought about by antagonism in the brain of the action of 5 HT. This, however, is by no means certain.

Quite a number of extracts of tissues have been shown to be pharmacologically active. The extraction procedure may produce or release these substances and it is, therefore, in most cases impossible to say whether they are active in physiological conditions. It is possible for example that free histamine appears in the skin only in allergic conditions or when damage has been done.

REFERENCES

Albright, F. & Reifenstein, E. C. (1948). *The Parathyroid Glands and Metabolic Bone Disease.* Baltimore: The Williams & Wilkins Co.

Bayliss, R. I. S. (1955). Factors influencing adrenocortical activity in health and disease. *Brit. med. J.* i, 495–501.

Bell, E. T. (1953). *Experimental Diabetes Mellitus.* Oxford: Blackwell.

Cannon, W. B. (1915). *Bodily Changes in Pain, Hunger, Fear and Rage.* New York: Appleton.

Dougherty, T. F. (1952). Effect of hormones on lymphatic tissue. *Physiol. Rev.* **32,** 379–401.

Gaddum, J. H. (Ed.) (1955). *Polypeptides which Stimulate Plain Muscle.* Edinburgh: Livingstone.

Greer, M. A. (1950). Nutrition and goiter. *Physiol. Rev.* **30,** 513–548.

Gross, J. & Pitt-Rivers, R. (1953). Recent knowledge of the biochemistry of the thyroid gland. *Vitam. & Horm.* **11,** 159–172.

Harington, C. R. (1933). *The Thyroid Gland, Its Chemistry and Physiology.* Oxford University Press.

Harris, G. W. (1953–54). Stress and thyroid activity. *Lect. Sci. Bas. Med.* **3,** 154–168.

Hechter, O. & Pincus, G. (1954). Genesis of adrenocortical secretion. *Physiol. Rev.* **34,** 459–496.

Hormones—A Survey of their Properties and Uses (1951). London: The Pharmaceutical Press.

Ingle, D. J. (1950). The biologic properties of cortisone: a Review. *J. clin. Endocr.* **10,** 1312–1354.

Means, J. H. (1948). *The Thyroid and Its Diseases.* 2nd ed. Philadelphis: Lippincott.

Means, J. H. (1954). *Lectures on the Thyroid.* London: Oxford University Press.

Page, I. H. (1954). Serotonin (5-hydroxytryptamine). *Physiol. Rev.* **34,** 563–588.

Pincus, G. (Ed.) (1948–56). *Recent Progress in Hormone Research.* New York: Academic Press Inc.

Pincus, G. & Thimann, K. V. (1948). *The Hormones.* Vol. 1, 1948; Vol. 2, 1950; Vol. 3, 1955. New York: Academic Press Inc.

Seyle, H. (1952). *The Story of the Adaptation Syndrome.* Montreal: Acta Inc.

Silva, M. & Rocha, E. (1955). *Histamine ; Its Role in Anaphylexis and Allergy.* Springfield: C. C. Thomas.

Symposium on the pathologic physiology of thyroid diseases (1956). *Amer. J. Med.* **20,** 651–744.

Vitamins and Hormones. Annual volumes. New York: Academic Press Inc.

Vogt, M. (1950). Regulation of secretion of cortical hormones. *Brit. med. J.* ii, 1242–1244.

Von Euler, U. S. (1951). Hormones of the sympathetic nervous system and adrenal medulla. *Brit. med. J.* i, 105–108.

Yoffey, J. M. (Ed.) (1953). *Colston Symposium on "The Suprarenal Cortex".* London: Butterworth.

CHAPTER 52

REPRODUCTION

THE essential difference between the two sexes is that the male possesses testes and the female ovaries. The gonads are formed from a ridge on the Wolffian body which is derived from the intermediate cell mass of the embryo. At a very early stage it is not possible to say whether a gonad is to develop into a testis or an ovary but differentiation has proceeded far enough in the human embryo by the seventh week to allow the sex to be recognized histologically. Associated with each developing gonad are two pairs of ducts, the Wolffian and the Müllerian. The former persist in the male as the vasa deferentia which lead, one on each side, from the testes to the urethra. In the female, only the Müllerian ducts, which arise as an invagination of the body cavity, persist to develop into the uterus; in the human subject, since the caudal ends of the Müllerian ducts become fused, the uterus has a single cavity.

The gonads have, through the hormones which they produce, a determining influence on growth and on behaviour.

THE TESTES

The testes develop during embryonic life in the abdominal cavity but each acquires an attachment, the gubernaculum testis, which connects it to that area of the skin which later forms the scrotum. As development proceeds the testes are carried to the bottom of the scrotum which they reach about the eighth month of intrauterine life. Sometimes this process of descent is delayed or incomplete; if the testes are retained in the abdominal cavity in extrauterine life (*cryptorchidism*) spermatozoa are not formed because the temperature of the abdominal cavity is too high. The cutaneous muscle of the scrotum, the *dartos muscle*, by contracting when the environment is cold and relaxing when it is warm so regulates the distance between the testis and the warm groin that the scrotal contents are kept about 2° C below that of the abdominal cavity. At this temperature spermatogenesis proceeds normally.

Each testis is an oval body about 5 cm long and 3 cm across, covered by the tunica albuginea, a bluish-white membrane of fibrous tissue, to which are attached fibrous septa dividing up the testis into some 300 compartments each containing about three convoluted *seminiferous tubules.* On cross-section these tubules, each of which is about 80 cm long, show a basement membrane on which three irregular layers of epithelial cells can be distinguished (Fig. 52, 4). The outermost layer of cubical cells or *spermatogonia* give rise by cell division (p. 1014) to the second layer, the *spermatocytes*, which are large cells with large nuclei. Two types of spermatocyte are found in this layer, namely the primary spermatocytes containing the somatic or diploid number of chromosomes (p. 1014) and the secondary spermatocytes formed from them by reduction division. The latter, containing only the haploid number of chromosomes, give rise by a further division to the spermatids, small polyhedral cells found in the third layer. By modification of their structure the spermatids become spermatozoa and during this process they cluster around the *supporting cells of Sertoli* which are large cells rich in glycogen and lipids projecting from the basement membrane. The mature spermatozoa leave the Sertoli cells and lie around the inner margin of the germinal epithelium where they can easily be recognised by their densely staining heads and long curved tails pointing into the lumen of the tubule. The high refractive index of ripe spermatozoa shows that they must contain nearly 50 per cent. of solid matter. Spermatozoa are first produced at puberty and spermatogenic activity is maintained into old age so that there is no definite end to reproductive life in the male. The germinal cells are easily damaged by ischæmia. In the rat occlusion of the testicular artery for one hour leads to irrevocable destruction of all spermatogenic tissue; only the cells of Sertoli survive and act as macrophages.

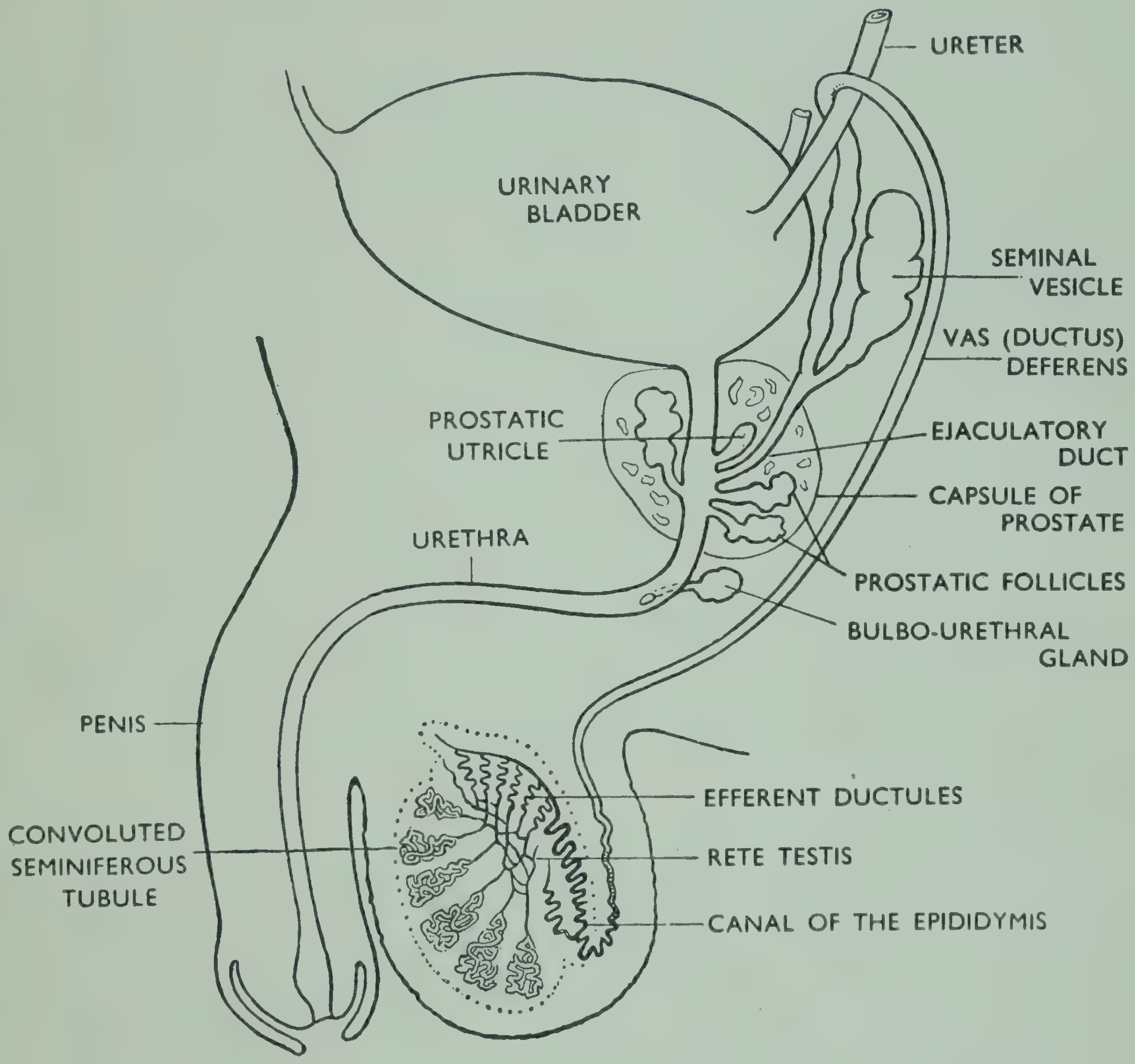

URETER
URINARY BLADDER
SEMINAL VESICLE
VAS (DUCTUS) DEFERENS
PROSTATIC UTRICLE
EJACULATORY DUCT
CAPSULE OF PROSTATE
URETHRA
PROSTATIC FOLLICLES
BULBO-URETHRAL GLAND
PENIS
EFFERENT DUCTULES
CONVOLUTED SEMINIFEROUS TUBULE
RETE TESTIS
CANAL OF THE EPIDIDYMIS

FIG. 52, 1

Diagram of the male sex organs.

From the seminiferous tubules (Fig. 52, 1) the spermatozoa pass into the *rete testis*, a series of channels in the fibrous stroma of the posterior part of the testis. About a dozen efferent ductules pass from the upper part of the rete testis into the single canal which by its convolutions makes up the *epididymis* where the spermatozoa are stored until ejaculation takes place. Production continues whether ejaculation occurs or not and spermatozoa not ejaculated are reabsorbed in the *vas deferens*. The canal of the epididymis is about six metres long and is continued into the vas deferens which rises up over the brim of the pelvis and passes down to the *prostate gland* at the base of the urinary bladder. Here it is joined by the duct of the *seminal vesicle* to form the short ejaculatory duct which passes through the prostate and enters the prostatic urethra. The prostate, surrounding the commencement of the urethra, is a glandular organ containing much muscle tissue and numerous follicles which communicate with the prostatic urethra by about a dozen small ducts. About 2·5 cm distally in the penile part of the urethra are the openings of the ducts of the two bulbo-urethral glands.

At coitus there is ejaculated into the vagina 2 to 4 ml. of semen containing several hundred million spermatozoa; the first portion ejaculated is derived mainly from the bulbourethral glands, this is followed by the prostatic secretion and finally by the spermatozoa suspended in vesicular secretion. Less than ten per cent. of the volume of the ejaculate is occupied by spermatozoa, the remainder being seminal plasma. The semen clots soon after ejaculation under enzymic influences but is liquified shortly after by the action of proteolytic enzymes; then the spermatozoa become fully motile. The composition of seminal plasma is rather variable from one individual to another and consecutive samples from the same individual may differ widely. Its pH is about 7·4 and it contains some 0·4 per cent. citric acid, 0·2 per cent. fructose, 0·01 per cent. ascorbic acid together with phosphorylcholine, inositol and various enzymes as described below. The seminal plasma acts as an activator and as a diluent for the spermatozoa which are tightly packed in the epididymis. The hyaluronidase of semen is carried by the spermatozoa but a certain amount leaks out into the plasma.

The prostatic secretion which forms about 20 per cent. of the volume of semen is colourless and slightly acid (pH 6·5) due to the presence of citric acid; it contains the proteolytic enzymes, fibrinolysin and fibrinogenase, and also acid phosphatase. The amount of acid phosphatase in the prostate is very small till puberty but it increases in adult life. Normally prostatic phosphatase does not pass into the blood stream but the acid phosphatase of male urine is due to admixture of prostatic secretion. In newly ejaculated semen this enzyme breaks down the phosphorylcholine to free choline and orthophosphate. The output of the enzyme in monkeys is increased by injection of androgens.

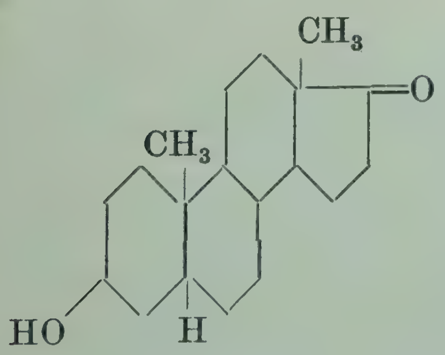
The seminal vesicles provide about 60 per cent. of the volume of the semen. Their secretion is yellow owing to the presence of flavins and is sometimes deeply pigmented. It contains a large amount of potassium and also ascorbic acid, ergothioneine, fructose and phosphorylcholine. This secretion is also under control of androgens.

Spermatozoa obtained from the epididymis can oxidize glucose rapidly but spermatozoa after ejaculation obtain energy from the seminal fluid and vaginal secretions by the anaerobic breakdown of the fructose to lactic acid which is oxidized, if oxygen is present, with liberation of energy. Fructose is therefore an important constituent of the medium in which spermatozoa are stored *in vitro* for use in artificial insemination. After castration, fructose, citric acid and phosphatase are not found in the seminal plasma.

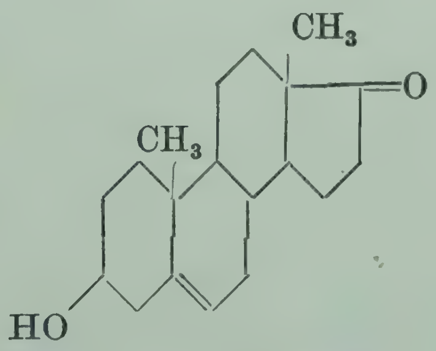
INTERNAL SECRETIONS OF THE TESTIS

Puberty occurs in the boy about fourteen to sixteen years of age but signs of pubescence, such as pubic hair, may be seen a year or more before this. The penis, the scrotum, the prostate and other accessory sex organs increase in size. The larynx enlarges and the pitch of the voice falls. The hair on the face, on the pubis and abdomen becomes thicker and grows more abundantly. At the same time there is rapid muscular development and an increased rate of growth with psychological changes and alteration of the pattern of behaviour. There is no doubt that these bodily and mental changes depend on the presence of the testes since they do not occur if these organs are removed before puberty. Castration, that is, removal of the testes, has been performed on domestic animals for thousands of years with the object of making the flesh more pleasant to eat or of making the animal more docile. Castration has also been performed on man for many centuries for ceremonial or other reasons. If the operation is carried out before puberty the eunuch usually becomes a tall man because of the prolonged period of growth of the long bones. Usually rather fat, he retains the treble voice of the boy, the penis remains small and hair does not grow over the face or abdomen in the typical masculine fashion; the common masculine form of baldness does not develop and sebaceous secretion is diminished. The skin of eunuchs does not tan on exposure to the sun unless male hormone has been administered. The prostate, seminal vesicles, and scrotum remain infantile and the thymus persists. Such an individual has no sexual desire and is both impotent and sterile. Castration after puberty produces little physical change except atrophy of the prostate and seminal vesicles; the histological changes in the pituitary are described on p. 970. Although psychological changes are common, sexual desire may be retained, and the individual though potent is sterile. Analogous physical changes are produced by experimental castration in other mammals. After removal of the anterior pituitary in animals or destruction of the anterior hypothalamus spermatogenesis fails and the testes atrophy.

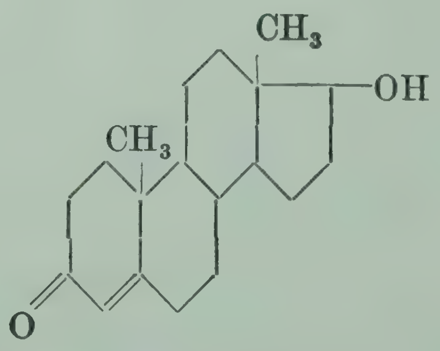
Ligature of the vasa deferentia is followed by atrophy of the seminiferous tubules of the testes but the general effects of the operation are quite different from those of castration since there is no atrophy of the prostate or seminal vesicles. It is assumed therefore that the male hormone is produced by groups of cells lying between the tubules, the *interstitial cells* of Leydig, which contain yellow pigment granules and possess a rich supply of nerves.



Androsterone



Dehydroisoandrosterone



Testosterone

The first important experiments in endocrinology were made by Berthold in 1849. If a cock is castrated the comb shrinks and the behaviour alters. Berthold found that cockerels which had their testes transplanted into a new position in the body maintained their normal male appearance and showed male sexual behaviour. Clearly, then, the characteristic male appearance and behaviour of the bird depends on the testis which is able to exercise its influence independently of the nervous system. Comb growth, the result of an increase in the mucoprotein content of the comb, is still one of the best quantitative biological tests for *androgens,* that is substances which bring about the development of male organs and maintain them in the absence of the testes. With the help of this test the chemical identity of the androgens has been elucidated (p. 27). They are steroids, the two best known being *androsterone* and *testosterone.* Androsterone was first isolated from urine in 1931 and a related substance, *dehydroandrosterone,* in 1934. Testosterone, first obtained in 1935 from the testes of bulls, is an alcohol which forms an ester with propionic acid, testosterone propionate, in which form the hormone is used in clinical medicine; it has to be given parenterally, that is to say subcutaneously or intra-muscularly because it is destroyed in the alimentary tract or more probably because after absorption it is destroyed in the liver. Methyl testosterone, however, may be given under the tongue since it is absorbed from the buccal mucosa. Testosterone is produced in the testis and may be regarded as the male hormone. Androsterone, on the other hand, while possessing androgenic activity is a urinary excretory product derived in part from the metabolism of testosterone; however, since the greater part of injected testosterone is destroyed in the body the amount of androsterone recovered in the urine is invariably less than the amount of testosterone administered. For this reason the androsterone excretion is not a satisfactory index of the exact amount of testosterone produced in the body; a high androsterone excretion, however, implies a high androgen production.

Androgens are excreted in the urine as water-soluble compounds such as sulphate esters which are biologically inactive. Acid hydrolysis yields the biologically active free androgen which is water-insoluble but fat-soluble. It is important for diagnostic purposes to be able to estimate the amount of androgens excreted and it has been found in the case of urine, though not necessarily in other tissues or body fluids, that there is a good correlation between the amount of androgenic material estimated by the comb growth or other biological test, and the amount of steroid with a keto grouping in position 17 of the ring system (p. 27). The 17-ketosteroids (17-oxosteroids), of which androsterone is the most important, can be estimated chemically by means of a characteristic red colour which they give in alkaline alcoholic solution with *m*-dinitrobenzene (Zimmerman reaction).

The amount of androgenic material excreted in the urine is given in Table 52, 2. The range of variation among healthy individuals is considerable; for example, according to one writer the urinary excretion of 17-ketosteroids in men is 8 to 20 mg. per day (mean 13·5) and in women 6 to 12 mg. per day (mean 9). Lower values are found in eunuchs and in cases of pituitary insufficiency. In normal human subjects the adrenal cortex is responsible for the production of some of the 17-ketosteroids which appear in the urine; it is not surprising, therefore, to find that overactivity of the cortex due to the presence in it of a tumour, or to stimulation of the cortex by the adrenocorticotrophic hormone of the pituitary, is characterized by a great increase in steroid excretion. The excretion of more than 35 mg. of 17-ketosteroids per day is suggestive of a suprarenal cortical tumour; values as high as 2000 mg. per day have been obtained in proved cases of this condition. In adrenocortical insufficiency (Addison's disease) the steroid excretion is reduced.

The testis produces, apparently in the tubules, a certain amount of female sex hormone, probably œstradiol, and small quantities are normally present in the urine of men. The functional significance of the œstrogen production in males is not known but it may act together with androgen in maintaining the fibro-muscular structure of the prostate and seminal vesicles. Since changes in the anterior pituitary occurring after castration cannot be reversed by injection of testosterone alone, other hormones (possible œstrogens) must be produced in the testes; these hormones may regulate the output of follicle stimulating hormone (FSH, p. 968) from the anterior pituitary. The androgens and œstrogens found in the urine after castration are probably derived from the suprarenal, since compounds having the biological effects of both male and female sex hormones are known to be produced in the suprarenal cortex.

TABLE 52, 2

	17-ketosteroids mg./24 hours	Œstrogen μg./24 hours	Pregnanediol mg./24 hours
Boys age 5	1	<1	—
Girls age 5	1	<1	—
Boys before puberty	7	1	—
Girls before puberty	7	2	—
Adult man	14	10	0·5
Adult woman	9	30	2
Senile man	5	10	—
Senile woman	3	5	—
Pregnant woman	9	100 to 20,000	10 to 100
Castrate man	7	1	—
Oöphorectomized woman	2	1	—
Female with Addison's disease (failure of adrenal cortex)	0	—	—
Male with Addison's disease	5	—	—

The amounts of 17-ketosteroids, œstrogen and pregnanediol excreted in the urine per day. These values are approximate only and wide variations from them are met with in individual cases. A dash in the table means that information could not be found in the literature.

Injection of androgens into castrated mammals prevents the atrophy of penis, scrotum, prostate, seminal vesicles, epididymis and bulbourethral glands which normally follows removal of the testes, and the increase in weight of the seminal vesicles of the castrated rat after androgen treatment is the basis of a commonly employed method of bioassay. Injection of androgens can also restore to normal the sexual behaviour of castrated mammals. Androgens can maintain normal spermatogenesis and prevent the testicular atrophy which follows on hypophysectomy in rats and in monkeys provided in the latter case that adequate amounts of the hormone are administered from the time of hypophysectomy.

In cases of delayed puberty in boys the administration of testosterone has produced hair in the typical male distribution and also growth of the larynx, and it has been successfully used to induce descent of the testis. Testosterone propionate given to a castrated man can produce erections of the penis and allow coitus. In castrated males with a genetic predisposition to it androgen therapy produces baldness and increases sebaceous secretion. (Baldness of some degree appears in nearly half of all adult males; androgens are only one of the inciting factors.) Repeated injections of testosterone in men have been reported to depress spermatogenesis but in a few cases after the injections cease spermatogenic activity is for a time greater than before.

It has been known for some time that androgens have an effect on protein anabolism in both man and animals. After an injection of testosterone propionate there is a retention of nitrogen and other tissue-forming materials such as potassium and phosphorus, which is greater if the supply of androgens from the patient's own testes is deficient. Testosterone propionate has been used successfully to induce a positive nitrogen balance and build up body tissue in cases of chronic undernutrition. Since certain urine extracts of unknown composition have a greater anabolic effect than would be expected from their androgen content, search is being made for other anabolic factors.

If the diet of experimental animals is inadequate in amount or deficient in vitamins, especially A, B or E, spermatogenesis is impaired and the accessory sex organs regress. In male rats such changes can be counteracted by administration of testosterone or anterior pituitary extracts. It is therefore supposed that inadequate feeding depresses the anterior pituitary so that the gonadotrophic stimulus to the testes declines and consequently their output of male hormone also diminishes.

THE FEMALE GENITAL ORGANS

The internal genital organs in the female consist of the ovaries, the uterine tubes, the uterus and the vagina (Fig. 52, 3). The *uterus* is a pear-shaped muscular organ about 7·5 cm long with a flattened triangular lumen, the upper part of which communicates through the *uterine tubes* with the peritoneal cavity, while the lower part is continuous through the narrow canal of the *cervix* with the *vagina* which passes to the exterior. The peritoneum covering the uterus is continuous with that lining the rest of the abdominal cavity. The walls, about 1 cm thick, are composed of smooth muscle fibres (the *myometrium*), among which is a considerable amount of fibrous tissue. There is very little elastic tissue in the body of the uterus but it is abundant in the cervical region. The mucous membrane or *endometrium* lining the interior, and continuous with the mucous membrane lining the uterine tubes and cervix, is covered by a columnar epithelium which dips down into the stroma of the endometrium to form simple tubular glands. The epithelium covering the interior of the tubes and uterus, with the exception of the lower part of the cervix, is provided with cilia which tend to move ova or secreted material towards the exterior. The columnar cells of the cervix give place, close to its external orifice, to the stratified squamous epithelium covering the vaginal portion of the cervix.

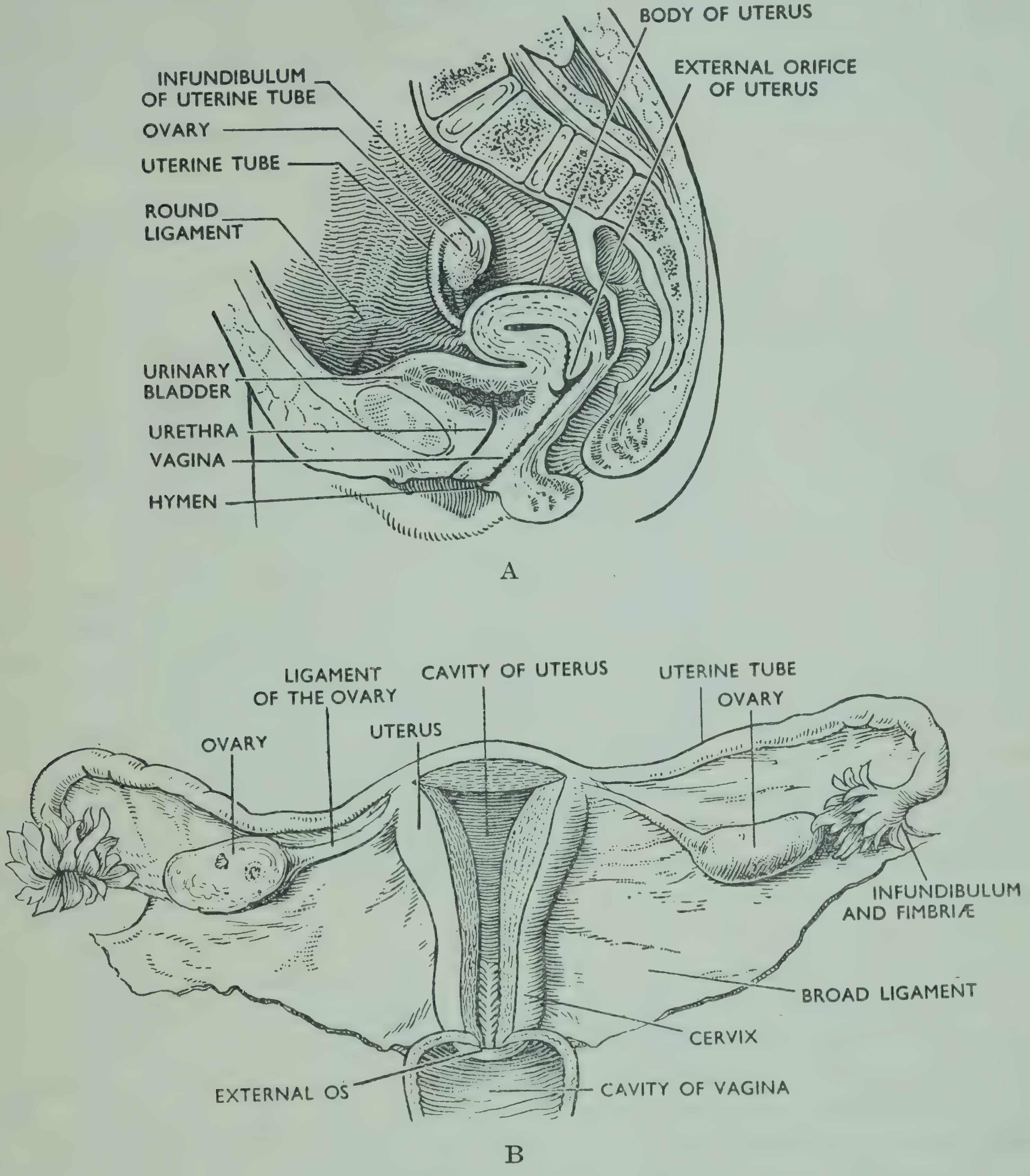

INFUNDIBULUM OF UTERINE TUBE
OVARY
UTERINE TUBE
ROUND LIGAMENT
URINARY BLADDER
URETHRA
VAGINA
HYMEN
BODY OF UTERUS
EXTERNAL ORIFICE OF UTERUS
A
LIGAMENT OF THE OVARY
CAVITY OF UTERUS
UTERINE TUBE
OVARY
UTERUS
OVARY
INFUNDIBULUM AND FIMBRIÆ
BROAD LIGAMENT
CERVIX
EXTERNAL OS
CAVITY OF VAGINA
B

Fig. 52, 3

Diagrams to show the main features of the female genital organs. A, sagittal section through the female pelvis. The uterus is shown in the so-called normal, anteflexed position; it is however retroflexed in 20 per cent. of normal women. B, the genital organs removed from the body (posterior aspect).

The Ovaries.—The ovaries in the adult female are oval bodies about 3 cm by 1·5 cm by 1 cm covered by a cubical epithelium called the *germinal epithelium.* The stroma or framework consisting of fibrous tissue with some muscle fibres is particularly dense below the epithelium where it forms the tunica albuginea. Immediately below this are numerous *vesicular ovarian* (*Graafian*) *follicles* in various stages of development or degeneration (Fig. 52, 5). There is considerable difference of opinion as to whether interstitial cells analogous with those of the testis are present in the adult human ovary; they can, however, be recognized in the ovary of the fœtus.

During fœtal life groups of cells grow down from the germinal epithelium, each group forming a little nest of cells known as the primordial follicle. At birth the ovary contains some 70,000 such follicles, the vast majority of which eventually disintegrate and disappear. From puberty onwards, however, a limited number mature in succession so that various stages of their development may be seen on one adult ovary. In each follicle one of the cells enlarges and can be recognized as a primitive *ovum.* It is very difficult to get direct evidence as to the time of maturing of the follicle destined to ovulate but in monkeys the enlargement of the follicle occupies only a few days before ovulation. Corner thinks that in man it is likely that the follicle begins to enlarge about the ninth day of the cycle and grows in four days from 2 mm to about 15 mm in diameter before rupturing. As the follicle begins to mature the cells investing the ovum multiply rapidly and a cavity filled with *liquor folliculi* appears among them giving the follicle the naked eye appearance of a clear vesicle on the surface of the ovary; the outer layer of investing cells is called the *stratum granulosum* and the layer around the ovum is the *cumulus ovaricus.* The maturing ovum is readily distinguished by its central position and relatively large size (0·13 mm). The ripe follicle is surrounded by a *tunica interna* consisting of spindle cells with numerous capillaries and a more compact *tunica externa* containing smooth muscle fibres. In the process of *ovulation* the follicle ruptures and the ovum, surrounded by cumulus cells, is set free into the abdominal cavity. The ovum then passes into a uterine tube. The empty space in the ruptured follicle is filled with blood and the clot is eventually replaced by large cells containing yellow pigment (*lutein cells*) which are mainly hypertrophied cells of the stratum granulosum although a few (*paralutein cells*) may be derived from the tunica interna. The resulting structure, known as a *corpus luteum,* persists till shortly before the next menstrual period if the ovum is not fertilized but if the woman becomes pregnant it continues to grow up to the third month of pregnancy, when it may be as large as 2·5 cm in diameter. In the later months it, diminishes in size so that at the end of pregnancy it is only 1 cm across.

Puberty Changes in the Girl

Puberty in the girl is signalized by the onset of periodic vaginal bleeding (the *menarche*) which occurs between 9 and 18 years; about half the girls of thirteen and a half years have begun to menstruate (Fig. 52, 7). There is no basis for the popular idea that the menarche is earlier in tropical than in temperate climates. Breast development may be seen more than two years before the menarche. During this stage described as pubescence

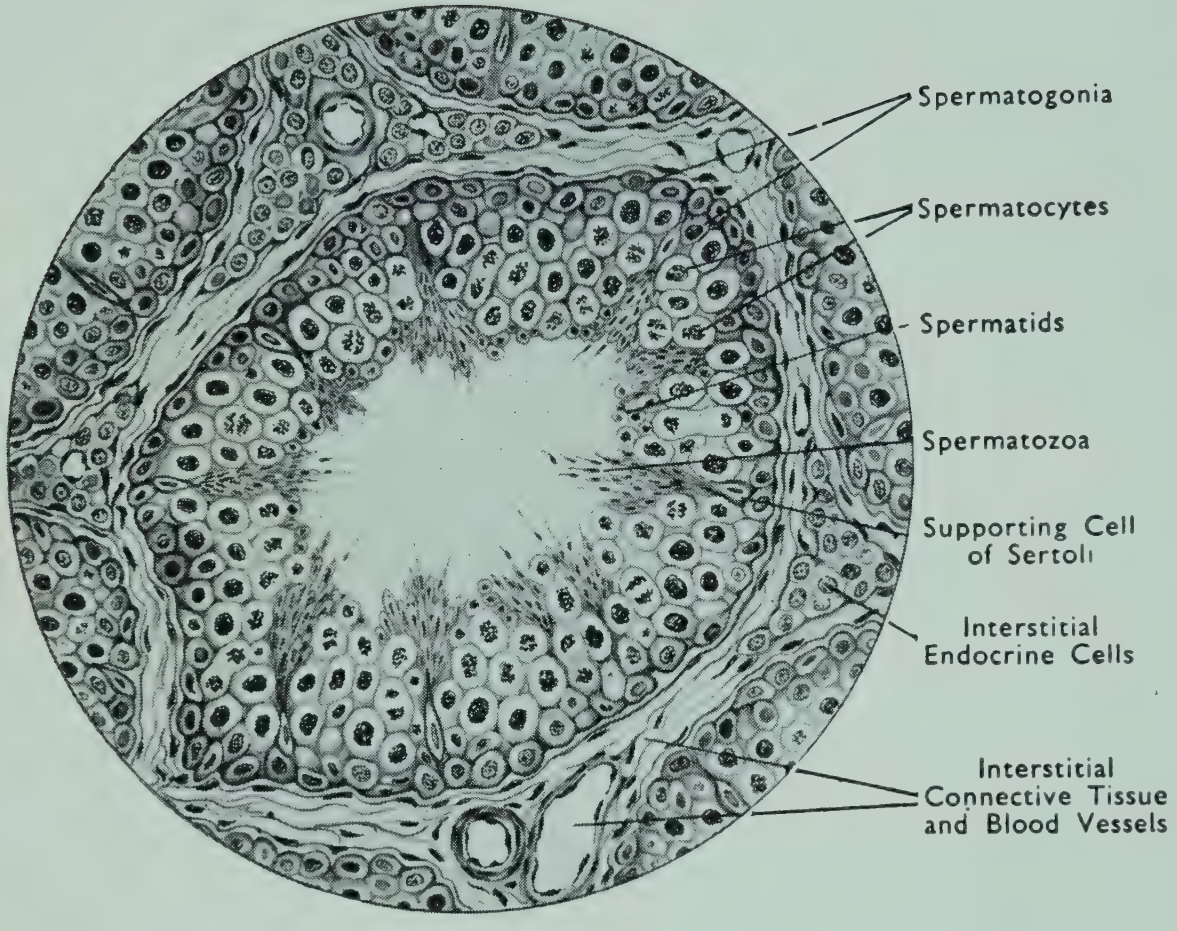

Spermatogonia
Spermatocytes
Spermatids
Spermatozoa
Supporting Cell of Sertoli
Interstitial Endocrine Cells
Interstitial Connective Tissue and Blood Vessels

FIG. 52, 4

Diagrammatic transverse section (× 185) through a convoluted seminiferous tubule of a cat (M. Gillison (1953), *A Histology of the Body Tissues.* Edinburgh: Livingstone).

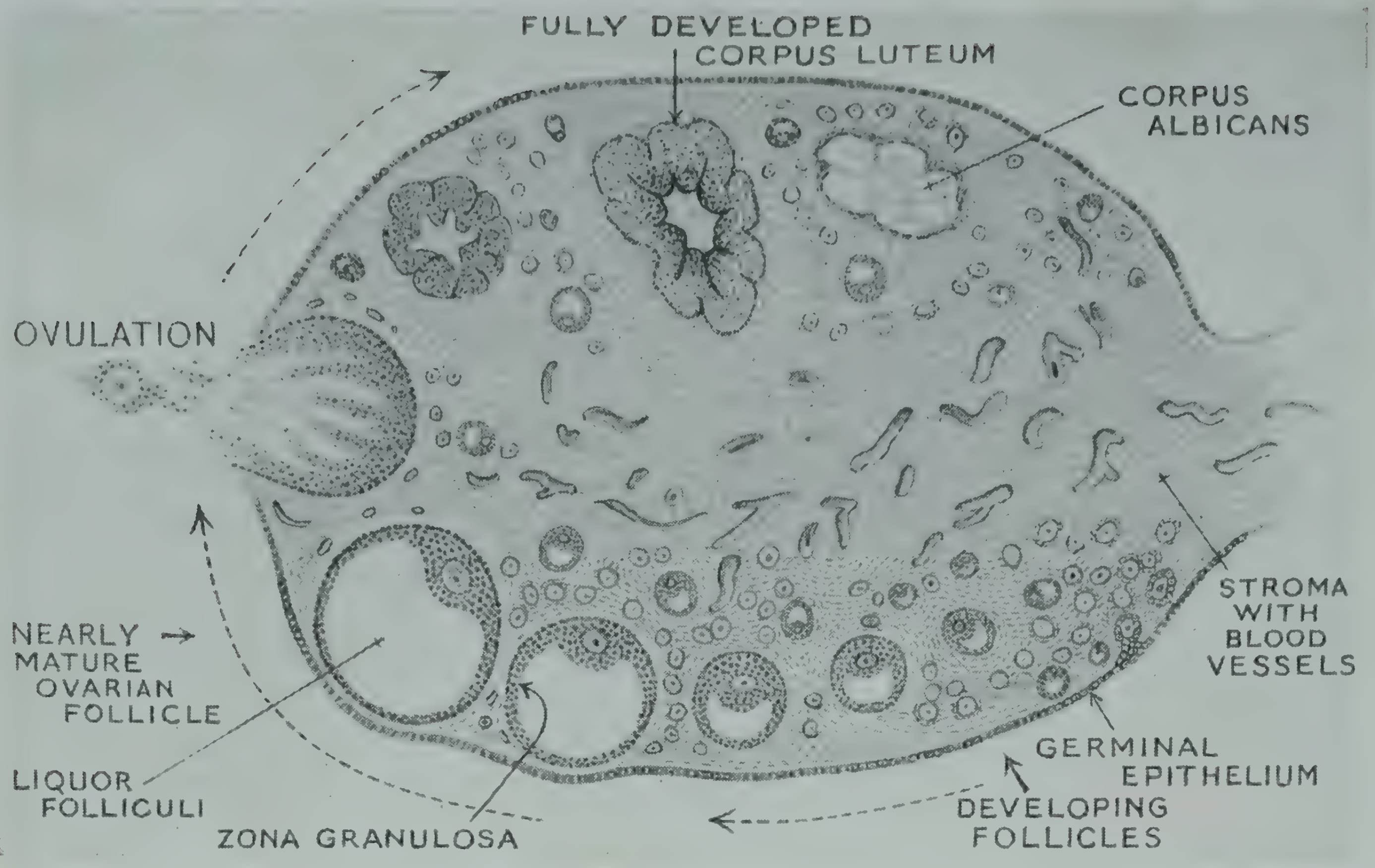

FULLY DEVELOPED CORPUS LUTEUM
CORPUS ALBICANS
OVULATION
STROMA WITH BLOOD VESSELS
NEARLY MATURE OVARIAN FOLLICLE
LIQUOR FOLLICULI
ZONA GRANULOSA
GERMINAL EPITHELIUM
DEVELOPING FOLLICLES

Fig. 52, 5

A diagrammatic representation of section made obliquely through an adult human ovary to show the stages in the development of the vesicular ovarian or Graafian follicle and the process of ovulation and finally the formation of a corpus luteum. The processes occur in the sequence indicated by the arrows.

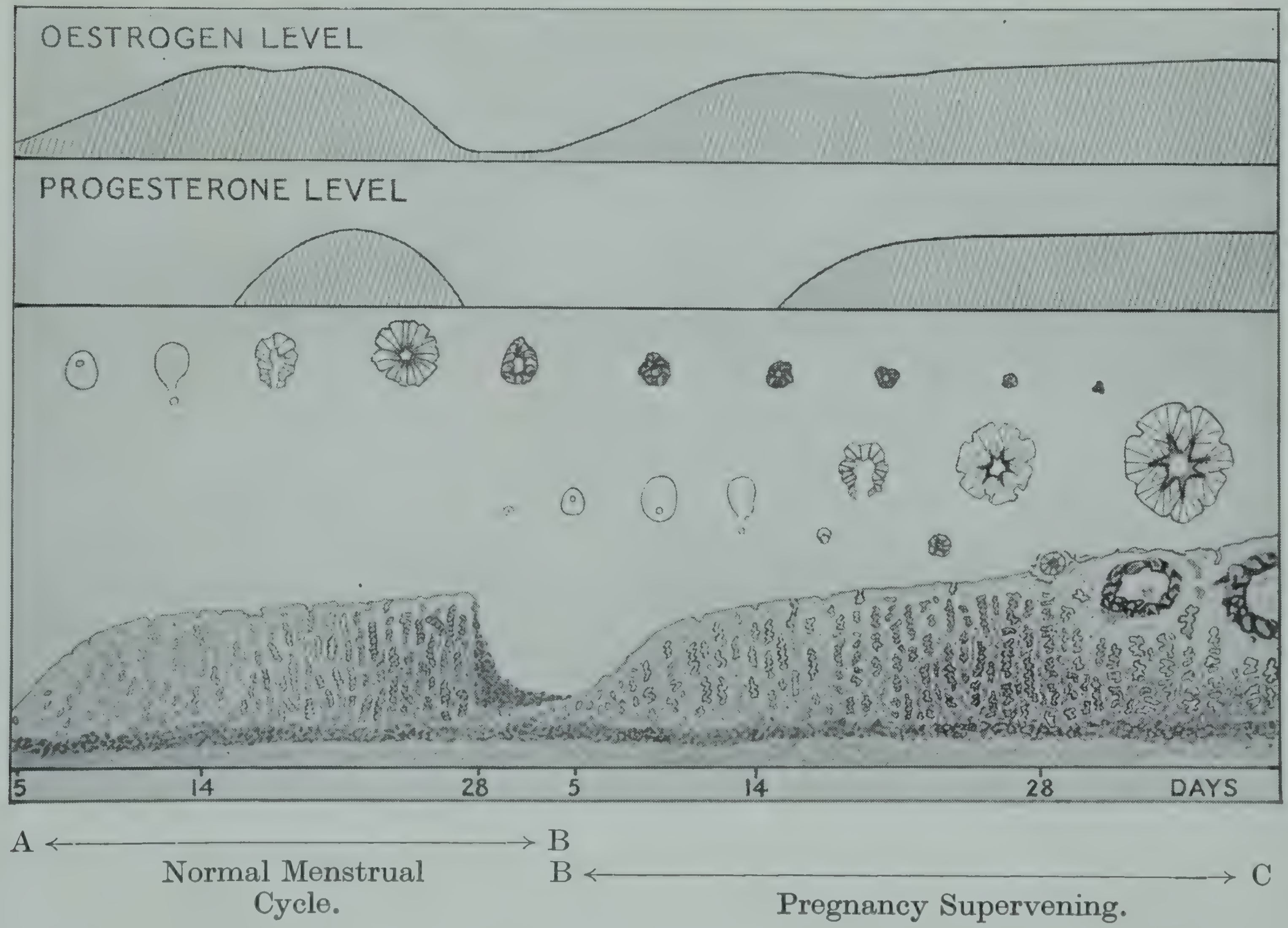

OESTROGEN LEVEL
PROGESTERONE LEVEL
1
2
3
5
14
28
5
14
28
DAYS
A ←——→ B
Normal Menstrual Cycle.
B ←——→ C
Pregnancy Supervening.

Fig. 52, 6

Scheme of the endometrial and ovarian events in a normal menstrual cycle and at the commencement of pregnancy. (*After* Schroeder.) It is now known that the progesterone level, as shown by the excretion of pregnanediol, does not fall to zero in the follicular phase of the cycle (see Fig. 52, 15).

1. Follicle and corpus luteum of cycle, A→B.
2. Follicle and corpus luteum of pregnancy, B→C.
3. Endometrium.

height increases rapidly, the rate of increase being less after the menarche. Puberty is accompanied by many changes in the reproductive organs as well as in the body generally, including the skeleton, which transforms the child into a woman who is able to bear children. The uterus and vulvar structures enlarge, pubic and axillary hair develops, and there is an increase in the size of the breasts, due mainly to deposition of fat but accompanied by some increase of the mammary tissue. Just before puberty in the girl the pelvis begins to grow rapidly, especially in the region of the sacro-iliac joints, so that its transverse diameter (breadth) increases relative to the antero-posterior diameter. The changes by which the infantile pelvis is converted to the adult female type with rounded inlet, shallow cavity and wide subpubic angle are said to occur within two years, after which there is little further change in the shape of the pelvis.

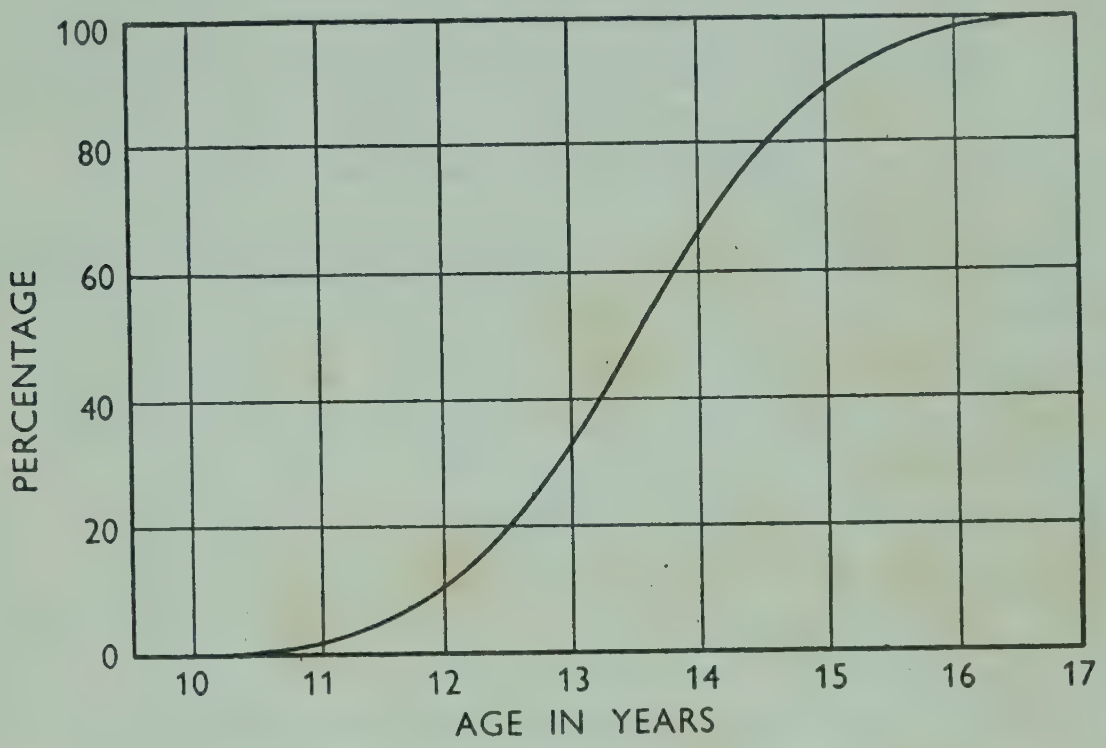


FIG. 52, 7

Graph showing percentage of girls who, at a given age, have attained the menarche (standard based on 2,590 girls living in the South of England) (D. C. Wilson & I. Sutherland (1950), *Brit. med. J.* ii, 862).

THE MENSTRUAL CYCLE

From the time of puberty the endometrium undergoes cyclical changes which result in bleeding from the uterus. This bleeding or *menstruation* lasts for four to six days and recurs every twenty-eight days approximately. In describing the menstrual cycle it is customary to date events in any one cycle from the first day on which bleeding occurs.

During the first half, or the *follicular phase* of the cycle, an ovarian follicle enlarges and comes to the surface of the ovary, ovulation occurring about the fourteenth day. The subsequent fourteen days in which the corpus luteum is active are referred to as the *luteal phase.* The corpus luteum degenerates just before the next menstruation (Fig. 52, 5, and Fig. 52, 6).

On the first three days or so of the menstrual cycle (Fig. 52, 9, D) the more superficial layers of the endometrium degenerate and are cast off with a certain amount of hæmorrhage, leaving the endometrium about 1 mm thick. Markee has studied the bleeding process by transplanting small fragments of endometrium into the anterior chamber of the eye of a monkey where they can easily be observed. In the follicular and luteal phases of the cycle the circulation through the spiral arteries supplying the endometrium is continuous. When the time of menstruation approaches there is a slowing of the circulation in the spiral arteries supplying the superficial parts of the endometrium and this is followed by a prolonged closure of these vessels which produces a blanching of the endometrium. After a time these arteries open again and blood escapes through their walls with the formation of subepithelial hæmatomata. The prolonged ischæmia leads to degeneration of the endometrium and hæmorrhage from capillaries, venules and arterioles. The cause of the vasoconstriction is not clear but it may be a consequence of the fall in the amount of ovarian hormones in the circulation at the end of the cycle.

The menstrual flow consists of blood mixed with mucus and at first contains numerous leucocytes. It does not clot but the reason for this is not known; anticoagulant substances manufactured by the endometrium may maintain the fluidity of the blood or it may be that after clotting in the uterus the blood is liquefied by enzymes present there. The volume of the blood lost varies greatly in different women but it is usually between 50 and 250 ml. The menstrual flow contains a toxic substance, since rats injected with this material die within twenty-four hours. Toxic material is also present in the endometrium just before menstruation but the nature and physiological function of the toxic factor are unknown.

At the end of menstruation the endometrium is regenerated (third to fourth day) from the deeper parts of the glands, next the myometrium, which are not affected by the vasoconstriction and are not cast off. During the *phase of proliferation*, occupying from the fifth to the fourteenth day, the endometrial glands progressively increase in length and tortuosity and the cells of the stroma hypertrophy so that at the end of this phase the endometrium is some 3 to 4 mm thick (Fig. 52, 9, A and B). During the next fourteen days the glands continue to grow, becoming more irregular in outline and much convoluted and showing secretory activity. About the end of this *secretory phase* the glands are dilated and contain mucus and glycogen; their walls are folded and the free edges of the cells are indistinct. The superficial cells of the stroma enlarge and since they possess more cytoplasm they give the stroma as a whole a more tightly packed appearance. In the deep layers the cells are also enlarged but the stroma has a more open and œdematous appearance (Fig. 52, 9, C). The basal part of the endometrium next to the muscle does not show œdema or hypertrophy. These premenstrual changes may be regarded as preparation for a fertilized ovum and when they are complete the endometrium is about 6 or 7 mm thick. If the ovum is not fertilized small collections of blood (hæmatomata) appear in the endometrium and the degenerative changes which result in menstruation begin again.

These histological details are derived from examination of small samples of endometrium obtained by inserting a curette through the cervix. The experienced histologist can tell from the microscopic appearances whether or not ovulation has occurred and can give with some degree of certainty the time of curettage relative to the day of ovulation. The precise day in any individual cycle cannot, however, be given since ovulation may occur any time between the sixth and twentieth day of the menstrual cycle. At the time of ovulation the cervix secretes watery mucus which presumably facilitates penetration of spermatozoa.

Bleeding normally takes place, as just described, from endometrium in the secretory phase but regular menstruation can occur even if the secretory stage is not reached and the endometrium, just before bleeding, is histologically still in the proliferative phase. Since the secretory phase of the endometrium is dependent on an active corpus luteum (p. 966) it can be presumed that in these cases ovulation has not occurred and the corpus luteum has not developed; bleeding of this kind is described as *anovular menstruation.* The corpus luteum is not essential for menstrual bleeding.

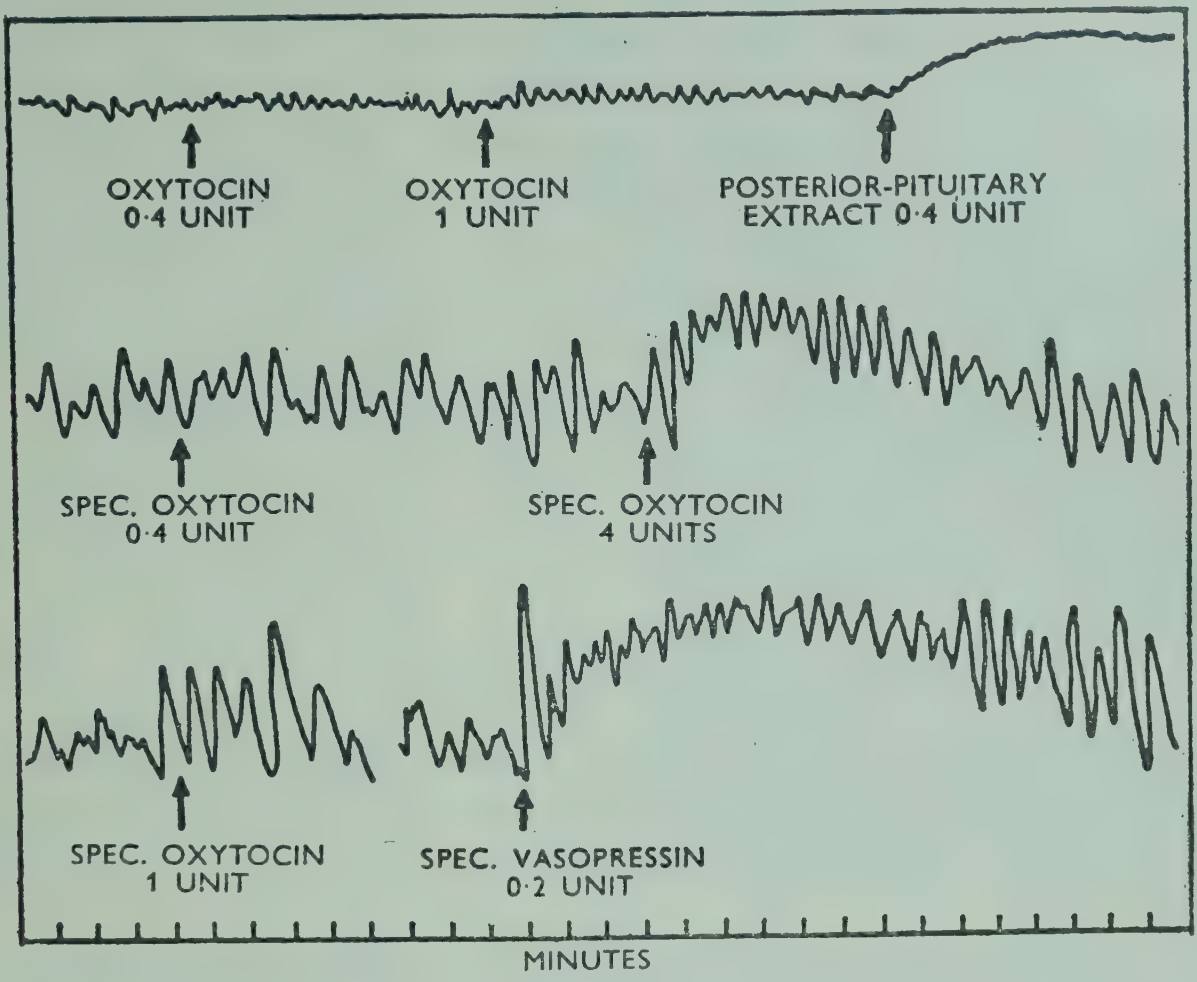


FIG. 52, 8

Records of uterine movements obtained by placing a bag within the uterus. An upward movement on the record indicates a contraction. Uppermost trace—the effect of oxytocin on the non-gravid human uterus on the 6th day of the menstrual cycle. Lower two traces show the effect of intravenous injection of specially purified samples of vasopressin and of oxytocin on non-gravid human uterus on the 26th day of the menstrual cycle. The specially purified oxytocin contained only 0·2 units of pressor activity for every 4 units of oxytocic activity. Since the same effect is observed with special oxytocin as with special vasopressin containing the same amount of pressor activity it can be said that the non-gravid human uterus does not respond to oxytocin. Note that spontaneous activity, before injection of drugs, is greater towards the end of the cycle (A. McLellan (1940), *Lancet*, i, 920).

The activity of the muscle of the non-pregnant human uterus can be recorded by placing in its cavity a rubber bag, filled with fluid under pressure, which is connected to a recording manometer. The uterine muscle shows, like the smooth muscle of the intestine, rhythmic contractions and relaxations, which, though present at all phases of the menstrual cycle, are usually more marked in the second half (Fig. 52, 8).

Although the uterus in other mammals in the non-pregnant state reacts to the oxytocic principle from the neurohypophysis (p. 1009) by contraction, the human myometrium in similar circumstances is stimulated by the vasopressor and not by the oxytocic principle (Fig. 52, 8).

In healthy women menstruation occurs on the average at intervals of twenty-eight days but there is a considerable range of variation among individuals. Further, in any one woman successive cycles may vary in length by one to two days. At the onset of menstruation, and also when the menopause is nearly reached, the cycles may be very irregular. The process is absent during pregnancy and ceases finally at the *menopause*, usually between forty-two and fifty-two years of age, when cyclical uterine and ovarian changes cease; uterus, breasts and external genitalia slowly atrophy and there is a gradual loss of sexual desire. This period of involutionary change, the *climacteric*, is often attended by vasomotor phenomena such as hot flushes and by giddiness, excessive sweating and temporary emotional disturbances. Similar uterine and vasomotor changes occur after an artificial menopause induced for therapeutic purposes by surgical removal or destructive irradiation of the ovaries.

ŒSTROUS CYCLE

The simplest œstrous cycle occurs in the rodents and it is therefore best to begin with a description of the cycle in the rat and mouse. The mature female of these species accepts the male only at the period of heat or *œstrus*, which lasts about a day and recurs every four or five days; ovulation occurs spontaneously at this time. The time between two periods of œstrus is called *diœstrus*. These cyclical changes do not occur after spaying, that is removal of both ovaries.

The vagina undergoes characteristic changes during the cycle and the discovery of these by Stockard and Papanicolaou in 1917 can be regarded as the starting-point of much of the modern work on sex hormones. These changes can be followed by microscopic examination of the cells in 'vaginal smears' made by passing a wire loop into the vagina and spreading the material so obtained on a glass slide. The original work was done on the guinea-pig but the technique has now been applied to man and many other animals. The detailed description varies somewhat in different species but in the rat at the height of œstrus the smear shows cornified squamous cells almost entirely, while at the end of œstrus large numbers of leucocytes appear (Fig. 52, 10). During diœstrus only a small amount of cellular material, chiefly leucocytes and epithelial cells, can be obtained. The uterus at the height of œstrus is enlarged and distended with fluid and the blood vessels are prominent; in diœstrus the vascularity declines and the distension passes off. The vaginal changes in woman are by no means so clear cut as in the rodents but cyclical alterations can be made out. Vaginal smears obtained in the first half of the human menstrual cycle show more and more cornified cells up to the fourteenth day, when leucocytes suddenly become abundant; in the second half of the cycle the proportion of cornified cells declines. At menstruation the vaginal smear is obscured by red blood cells.

A female mouse kept in isolation from the male shows regular cycles of five days' duration, indicated in the upper part of Fig. 52, 11, in which there

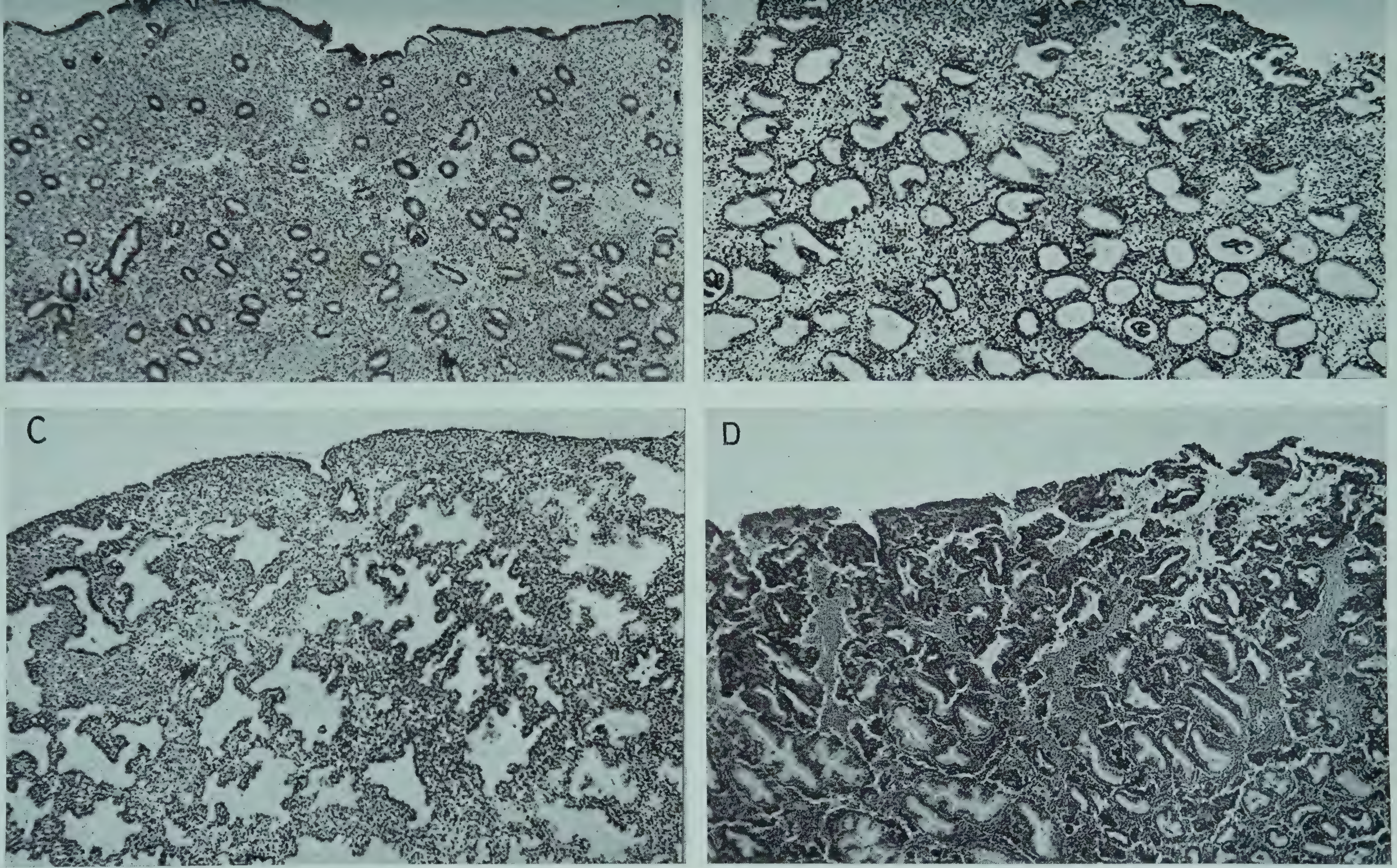
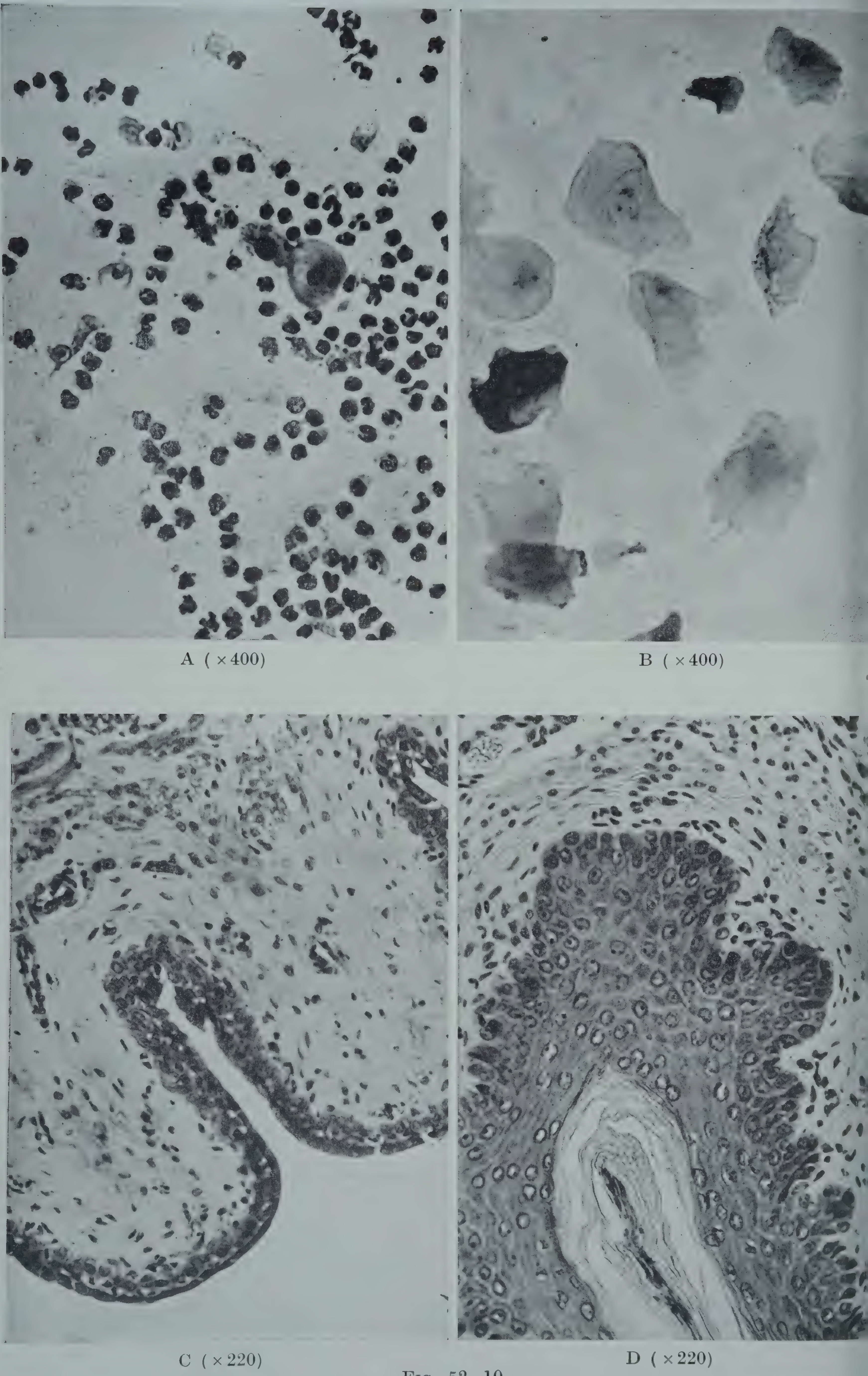


Fig. 52, 9

Human endometrium at different stages of the menstrual cycle. (×45). Stained with hæmatoxylin and eosin.

A. Reparative or early proliferative phase: 6th to 7th day of cycle.
B. Late proliferative phase: 12th to 13th day of cycle.
C. Secretory phase: 23rd to 24th day of cycle.
D Commencing menstruation.

(By courtesy of W. W. Park.)



A (×400)

B (×400)

C (×220)

D (×220)

Fig. 52 10

are no active corpora lutea. If, however, she is mated with a sterile male the corpora lutea formed in the ovary actively secrete their hormone and the next œstrus is delayed. This period of eleven days or *pseudopregnancy* is followed by œstrous cycles as before.

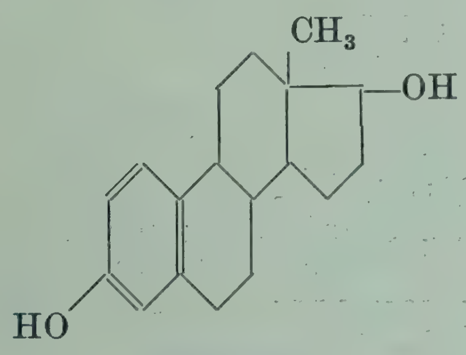
The bitch comes into heat twice in the year, the intervening quiescent period being called *anœstrus*. The week or so preceding œstrus, that is *pro-œstrus*, is characterized by growth of ovarian follicles and of the endometrium and vulval bleeding. Œstrus lasts for a week and during this time the male is accepted. Ovulation occurs during œstrus which is marked by vaginal bleeding due to further hæmorrhage from the endometrium. Active corpora lutea form in the ovary without any mating stimulus and the animal is said to be pseudopregnant. Two months later, at the end of pseudopregnancy, the endometrium breaks down again with vaginal bleeding and anœstrus follows.

In woman menstrual cycles succeed one another regularly so that there is no period of anœstrus but each cycle bears some similarity to the changes of pro-œstrus and pseudopregnancy in the bitch, except that there is no recognizable period of heat or œstrus and there is no bleeding at the time of ovulation. Here again no mating stimulus is required to produce the premenstrual or pseudopregnancy changes.

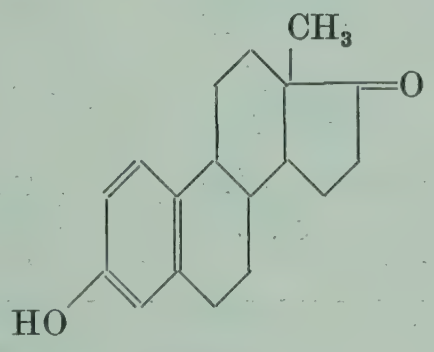
The sex cycles of the mouse, bitch and woman are compared diagrammatically in Fig. 52, 11.

OVARIAN HORMONES

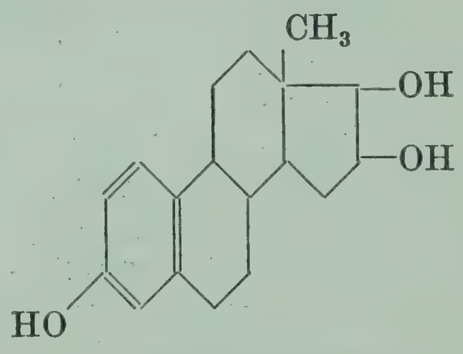
Œstrogens.—After spaying the uterus of the mouse and rat shrinks and the vaginal epithelium is reduced to a few layers of cells (Fig. 52, 10 C). By showing that the injection of extracts of liquor folliculi into spayed mice produced vaginal cornification (Fig. 52, 10 D) Allen and Doisy in 1923 found a method for testing œstrogenic activity. All substances producing œstrus and vaginal cornification in spayed mice or rats are said to be œstrogenic substances or *œstrogens*. As a result of the application of this test *œstrone* was isolated from urine in 1929 and *œstriol* by Marrian in 1930,



Œstradiol



Œstrone



Œstriol

Fig. 52, 10

Cellular changes in the rat vagina.

is a vaginal smear from a rat spayed fourteen days previously. It contains degenerating leucocytes and a few cells from the vaginal epithelium.

is a transverse section of the vagina of the same animal. The epithelium is quite thin.

is a vaginal smear from a spayed rat given a large dose of œstradiol dipropionate a week before. The cells are large non-nucleated squames.

is a transverse section of the vagina of the same animal. Note how the œstrogen treatment has increased the thickness of the vaginal epithelium. There is considerable cornification of the superficial layers in D. When these cells are cast off they appear as squames (B).

(*By courtesy of W. W. Park.*)

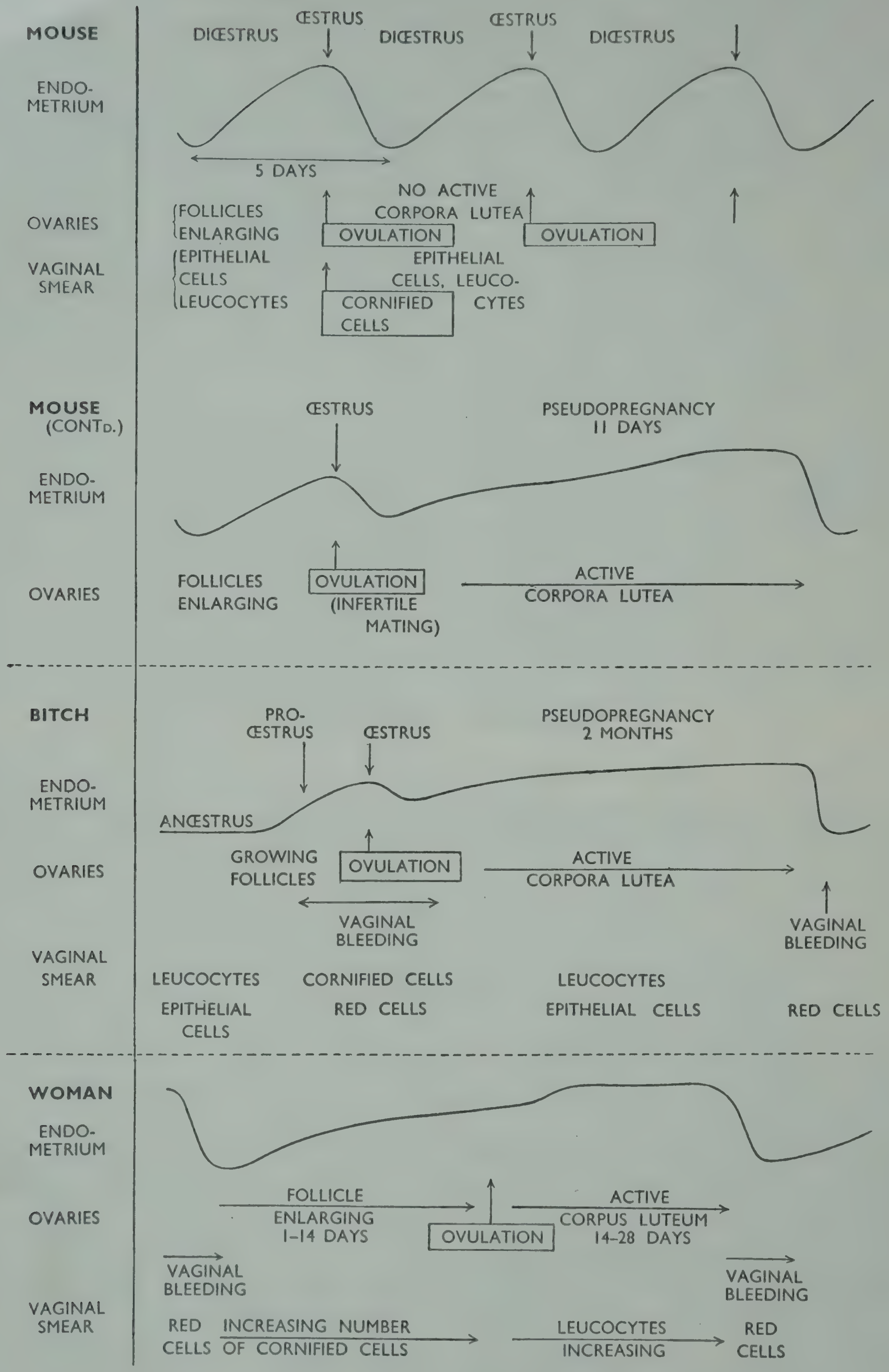

MOUSE
DIŒSTRUS
ŒSTRUS
DIŒSTRUS
ŒSTRUS
DIŒSTRUS
ENDO-METRIUM
5 DAYS
OVARIES
FOLLICLES ENLARGING
NO ACTIVE CORPORA LUTEA
OVULATION
OVULATION
VAGINAL SMEAR
EPITHELIAL CELLS LEUCOCYTES
CORNIFIED CELLS
EPITHELIAL CELLS, LEUCO-CYTES
MOUSE (CONTD.)
ŒSTRUS
PSEUDOPREGNANCY 11 DAYS
ENDO-METRIUM
OVARIES
FOLLICLES ENLARGING
OVULATION (INFERTILE MATING)
ACTIVE CORPORA LUTEA
BITCH
PRO-ŒSTRUS
ŒSTRUS
PSEUDOPREGNANCY 2 MONTHS
ENDO-METRIUM
ANŒSTRUS
OVARIES
GROWING FOLLICLES
OVULATION
ACTIVE CORPORA LUTEA
VAGINAL BLEEDING
VAGINAL BLEEDING
VAGINAL SMEAR
LEUCOCYTES EPITHELIAL CELLS
CORNIFIED CELLS RED CELLS
LEUCOCYTES EPITHELIAL CELLS
RED CELLS
WOMAN
ENDO-METRIUM
OVARIES
FOLLICLE ENLARGING 1–14 DAYS
OVULATION
ACTIVE CORPUS LUTEUM 14–28 DAYS
VAGINAL BLEEDING
VAGINAL BLEEDING
VAGINAL SMEAR
RED CELLS
INCREASING NUMBER OF CORNIFIED CELLS
LEUCOCYTES INCREASING
RED CELLS

FIG. 52, 11

The sex cycles of the mouse, bitch and woman diagrammatically compared.

and the presence in ovarian tissue of *œstradiol* was proved in 1935. Œstradiol, the hormone produced in the ovary, undergoes slight structural alterations in the body and is excreted as œstrone and œstriol.

These three œstrogens are all steroids, the structure of which has already been mentioned on p. 27. The naturally occurring œstrogens are but slightly active when given by mouth and must therefore be given parenterally. Ethinyl œstradiol, however, is effective by oral administration. The action of injected œstradiol is short but more prolonged effects can be obtained by administering œstradiol benzoate or dipropionate, or by intramuscular implantation of a tablet of œstrone which is slowly absorbed. Prolonged œstrogenic effects are, however, more simply and cheaply achieved by the oral administration of synthetic œstrogens (p. 964).

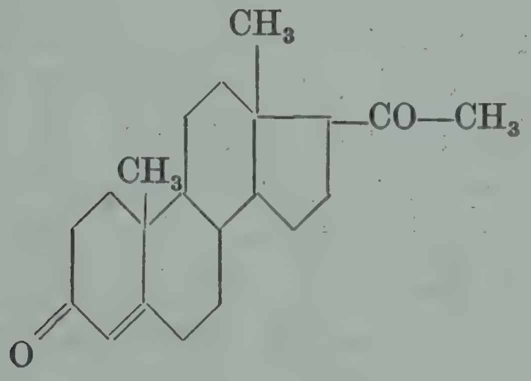
Not only can œstrogens produce the vaginal changes of œstrus but also, if given in sufficient amount, the enlargement and distension of the uterus which occur naturally at œstrus in rodents. If spayed mice are given still larger doses of œstrogens they will mate, but other animals, for example the guinea-pig, must be given progesterone along with œstrogen to produce mating behaviour. In higher animals, certainly in man, mating behaviour is not, however, determined by the amount of œstrogen in circulation; it is mainly under the control of the nervous system.

In most animals injections of œstrogen increase the spontaneous activity of the muscle (Fig. 52, 13) both of the body of the uterus and of the uterine tubes. In some species the sensitivity of the uterus to oxytocin, the hormone of the neurohypophysis which stimulates the uterus, is increased by administration of œstrogen but the effect on the response to oxytocin of the human uterus does not appear to have been measured. The non-pregnant human uterus is in any case not responsive to oxytocin (p. 959) but only to vasopressin. These differences, and others which are mentioned elsewhere in this chapter, make it clear that there are marked species variations in the pattern of reproductive physiology.

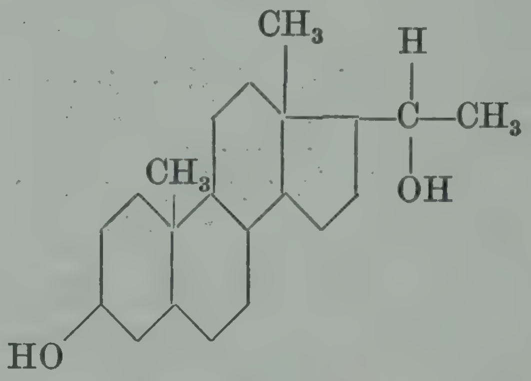
After removal of the ovaries from the rabbit the concentration of actomyosin (AM) in the uterus falls to one-fifth of that in the œstrous uterus; the concentration of AM is restored by giving œstrogens and so is the content of energy-rich phosphate. At the end of pregnancy the AM concentration in the uterus is even higher; perhaps this is due to the high level of œstrogens in the circulation at this time.

A certain amount of androgen is also secreted by the ovary.

Hormone of the Corpus Luteum.—It has long been suspected that the corpus luteum produces a hormone since, when actively secreting corpora lutea are present in the ovary, œstrus is absent and the endometrium shows characteristic changes. In the cow œstrus and ovulation may be suppressed when there is a persistent corpus luteum but if this is expressed from the ovary by manipulation per rectum the animal comes into heat again and can be mated. Furthermore in the presence of corpora lutea the endometrium of the guinea-pig, rabbit and monkey reacts to trauma, produced experimentally by inserting a piece of silk through the uterus, by the *decidual reaction* in which the connective tissue cells at the site of the injury enlarge. This reaction is also found in the early stages of implantation of a fertilized ovum (p. 976). The most useful method, however, for testing ovarian extracts for luteal activity depends on their action on the endometrium of the rabbit. The animal is first spayed and treated with œstradiol to produce the endometrial appearance shown in Fig. 52, 12, B. If the animal is then injected with active extracts of corpus luteum, the endometrial picture changes to the fronded appearance (Fig. 52, 12, C) known as *progestational proliferation*, which has been made by McPhail the basis of a method of biological assay. With the help of this test crude extracts were prepared in 1929, and in 1934 the pure hormone *progesterone* was isolated and its constitution determined. It is a steroid substance with a short side chain of two carbon atoms in position 17. In the liver this compound is reduced, by addition of six hydrogen atoms, to *pregnanediol* and conjugated with glucuronic acid to be excreted in the urine as a biologically inactive glucuronide. About 10 to 15 per cent. of injected progesterone can be recovered in the urine and about 5 per cent. in the bile as conjugated pregnanediol. The fate of the remaining 80 per cent. is not known. The concentration of pregnanediol can now be estimated by chemical methods.



Progesterone



Pregnanediol

Unless very large doses are used progesterone has no action on a tissue unless it has first been prepared ('primed') by œstrogen and, since the actions of progesterone and œstrogens are in some respects antagonistic, a delicately adjusted quantitative relationship between them is necessary to produce the maximum effect. Thus the changes produced in the endometrium by injections of progesterone in the rabbit or monkey are reduced or even completely prevented by large doses of œstrogen; the maximal progestational effect cannot, however, be obtained below a certain level of œstrogenic stimulation. Progesterone is inactive when given by mouth.

It is usually stated, on the basis of early experiments on the rabbit by Knaus, that progesterone reduces uterine activity and makes the uterus unresponsive to oxytocin. Later experiments have, however, not confirmed this for when the activity of the rabbit uterus is recorded by making successive X-ray photographs of silver-wire stitches tied alongside it there is no reduction of spontaneous activity after progesterone treatment (Fig. 52, 13). Moreover, in other species, although the rhythm of the spontaneous uterine activity may be altered, it is not reduced by the luteal hormone nor is the response to oxytocin abolished.

It seems that the function of progesterone, in conjunction with œstrogenic hormones, is to prepare the uterus for the reception of a fertilized ovum and to maintain the conditions necessary for its further growth.

Synthetic Œstrogens.—Large numbers of œstrogenic compounds having little or no chemical resemblance to the naturally occurring sex hormones have been synthesized. Stilbœstrol and hexœstrol are the best known examples.

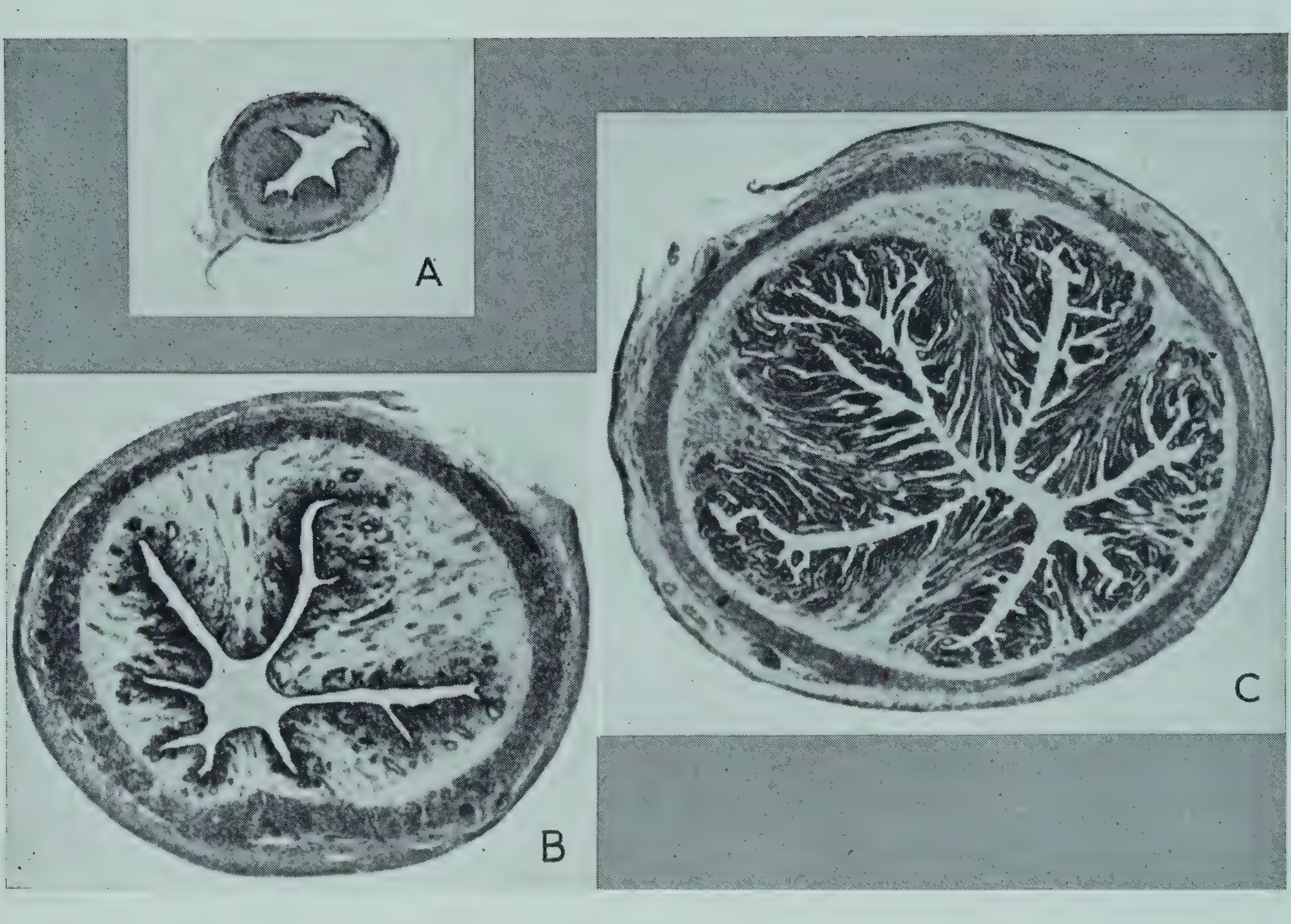


FIG. 52, 12

A. A transverse section of the uterine horn of a rabbit which had been ovariectomized some days previously. B. The same after treatment with œstrone. C. The same after treatment with œstrone and progesterone (progestational proliferation). Magnification in all cases × 17 (M. K. McPhail (1934), *J. Physiol.* **83,** 148).

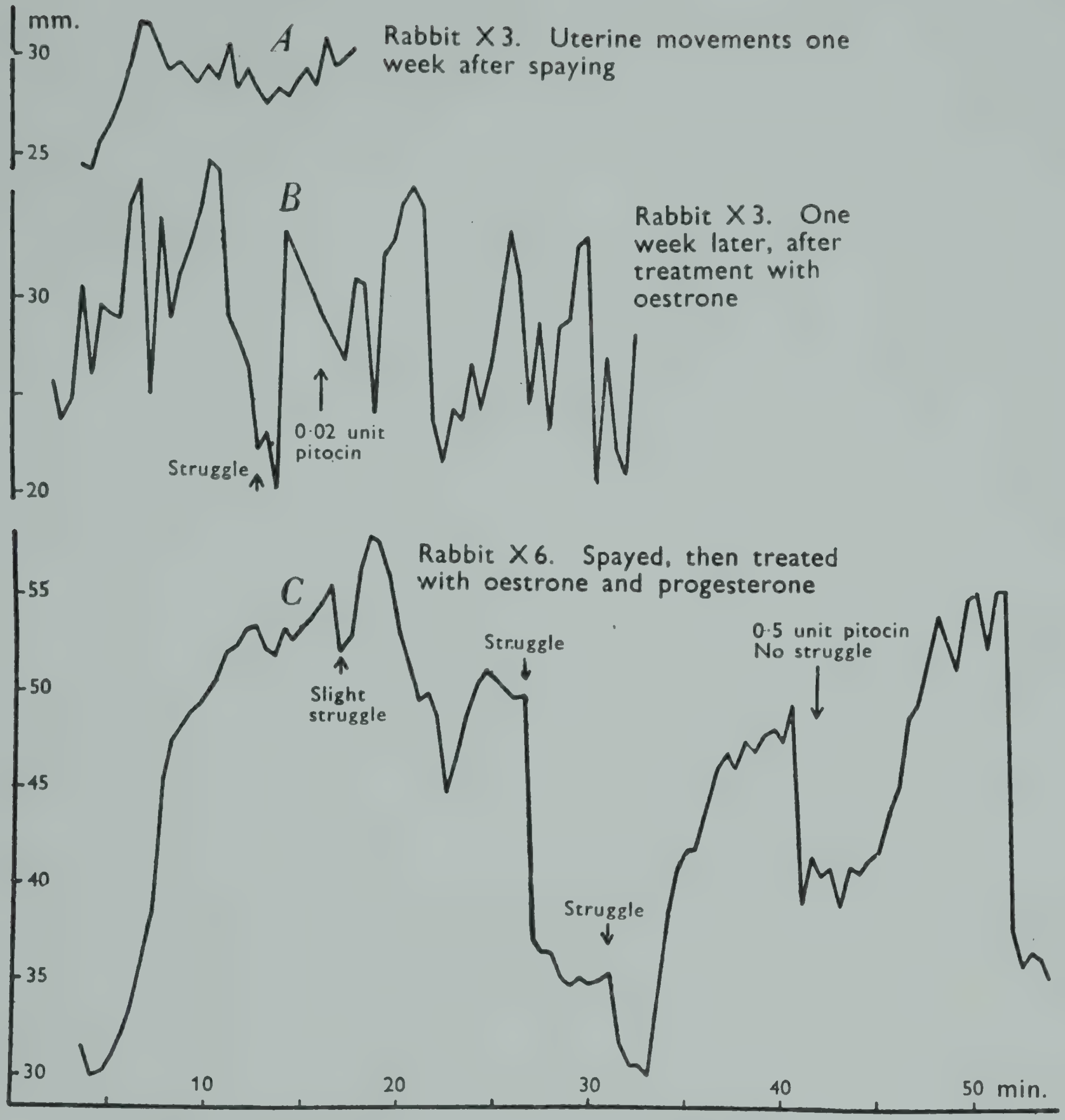
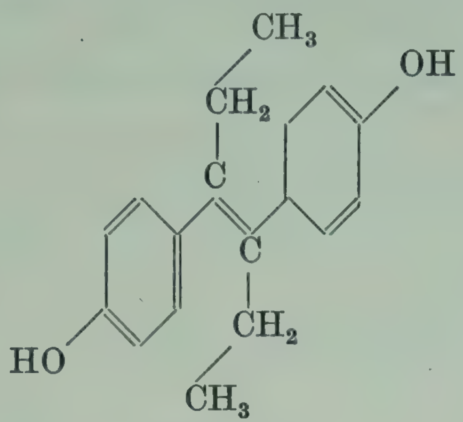


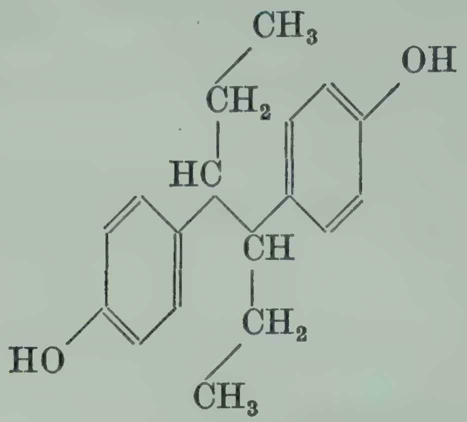
Fig. 52, 13

Silver-wire stitches were inserted alongside the margin of the uterus of the rabbit and the animal was oöphorectomized. Some weeks later a series of X-ray photographs of the abdomen was made and the distances between the stitches were measured. The sum of the distances was taken as the length of the horn at the instant when the photograph was taken and this is plotted on the ordinate (upper curve). Two further series of photographs were made after œstrone treatment (middle curve) and after progesterone treatment (lowest curve). Note the increase in the average length of the uterus with hormone treatment and the alteration in the rhythmic activity (G. H. Bell (1939), *J. Physiol.* **95,** 8 *P*).

For use in treatment these compounds have the great advantage that they are easily and cheaply prepared in bulk and are active when given by mouth. Œstrogens, both synthetic and natural, have proved of value in the palliative treatment of cancer of the prostate and cancer of the breast.



Stilbœstrol
(Diethylstilbœstrol)



Hexœstrol

Progesterone is unique in its progestational activity. There are very few other compounds which show any activity; ethynyl testosterone, the best known synthetic substance, has the advantage of being active by mouth.

All the steroid hormones, œstrogens, androgens and cortical hormones, are inactivated or destroyed to varying extents in the liver. Most information is available about the destruction of œstrogens as they are more easily estimated. About 5 per cent. of œstrone injected into a man is recovered in his urine. Œstrogens incubated with liver slices or perfused through the liver are partly destroyed. Castrated female rats with œstrogen pellets implanted in the spleen do not go into heat unless the spleen is transplanted from the portal to the systemic circulation. The destruction or inactivation in the liver does not occur in animals deficient in thiamine, riboflavin and methionine and thus implantation of œstrogens in the spleen of castrated animals deficient in these vitamins produces œstrus. It is, therefore, easy to see why impairment of hepatic functions in man by malnutrition, hepatic poisons or cirrhosis of the liver may show itself by excessive menstrual loss of blood in women or enlargement of the breasts or testicular atrophy in men. Palmar erythema and cutaneous spider nævi sometimes seen in chronic liver failure in man have been attributed to excessive action of œstrogens.

THE HORMONAL CONTROL OF THE MENSTRUAL CYCLE

The study of the changes occurring in the genital tract of women who have been given hormone treatment after removal of the ovaries provides important clues to the parts played by the various hormones in the human menstrual cycle. After öophorectomy the vagina atrophies and the vaginal smear consists almost entirely of leucocytes; at the same time the uterus atrophies and the endometrium becomes very thin. Treatment with œstrogens causes the walls of the vagina to become cornified so that the smear contains squamous cells and the vaginal secretion returns to its normal pH (about 4·5). The uterine muscle hypertrophies and the endometrium proliferates to reach eventually the stage characteristic of the middle of a normal menstrual cycle. At the same time the glands of the cervix uteri begin to secrete again. If daily injections are stopped or if the dosage is suddenly reduced the uterine mucosa becomes œdematous, and about five days later there is bleeding from the endometrium. This is called *withdrawal bleeding.* If, however, treatment is continued when the endometrium is showing proliferative changes, and progesterone as well as œstrogen is injected in suitable dosage a typical premenstrual endometrium can be produced. This treatment increases the vascularity of the endometrium and the spiral arteries have been seen growing towards the surface of the endometrium in an endometrial transplant in the anterior chamber of the eye of a monkey injected with these hormones. Cessation of hormonal treatment is followed as before by bleeding but in this case it takes place in a premenstrual endometrium just as in the normal menstrual cycle.

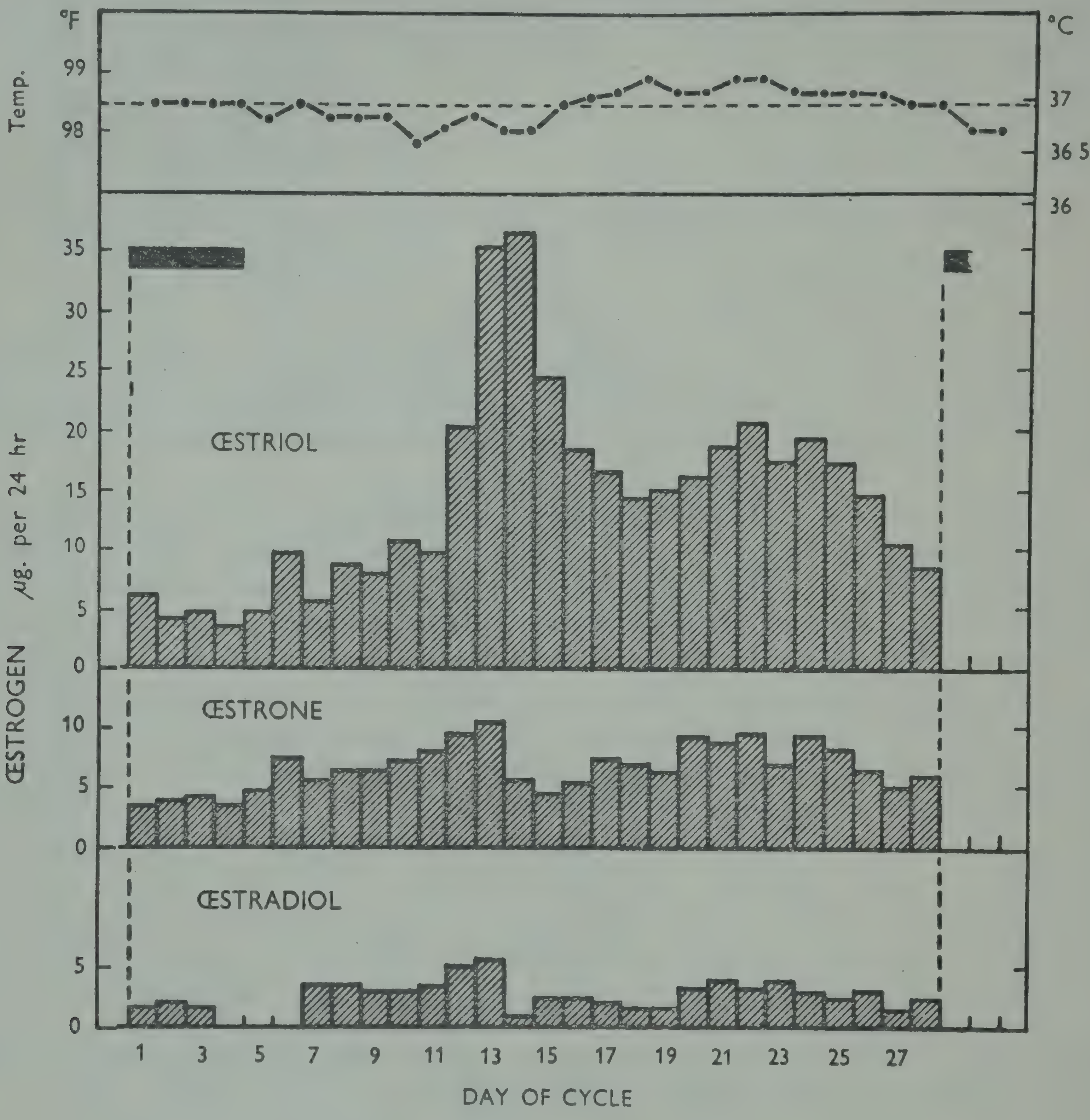


FIG. 52, 14

Urinary excretion of œstrogens throughout a normal menstrual cycle. The top graph shows the waking body temperature. The black bars indicate menstruation. (J. B. Brown (1955), *Lancet*, i, 320.)

Since the uterine changes of the menstrual cycle can be imitated by injection of œstradiol and progesterone, it is a fair presumption that the ovarian hormones are responsible for these changes in the normal menstrual cycle; œstradiol produced in the first half of the cycle would bring about the changes of the follicular phase while œstradiol and progesterone together would be responsible for the changes occurring in the luteal phase. This view is supported by the study of the urinary excretion of the end-products of the metabolism of the ovarian hormones. Fig. 52, 14, shows that œstrogens are excreted throughout the cycle and that there are two peaks of excretion, one about the time of ovulation and the other during the luteal phase. Observations on pregnanediol excretion (Fig. 52, 15) show that in the proliferative phase the daily excretion is 1·3 mg. on the average (minimum 0·9 mg.) and in the luteal phase 2·5 mg. (maximum 4·2 mg.). The actual production of progesterone can be assumed to be about 15 mg. per day. The rise in output of pregnanediol usually begins before the ovulatory rise of temperature and the output begins to fall before the onset of the period. Males and postmenopausal women excrete about 0·7 mg. per 24 hr; this excretion is abolished by adrenalectomy. Since in the first half of the cycle in women much more than this is excreted, it is likely that the functioning ovary, even in the follicular phase, produces some progesterone and in fact progesterone has been found in the fluid from mature follicles in various species and even in the blood draining from

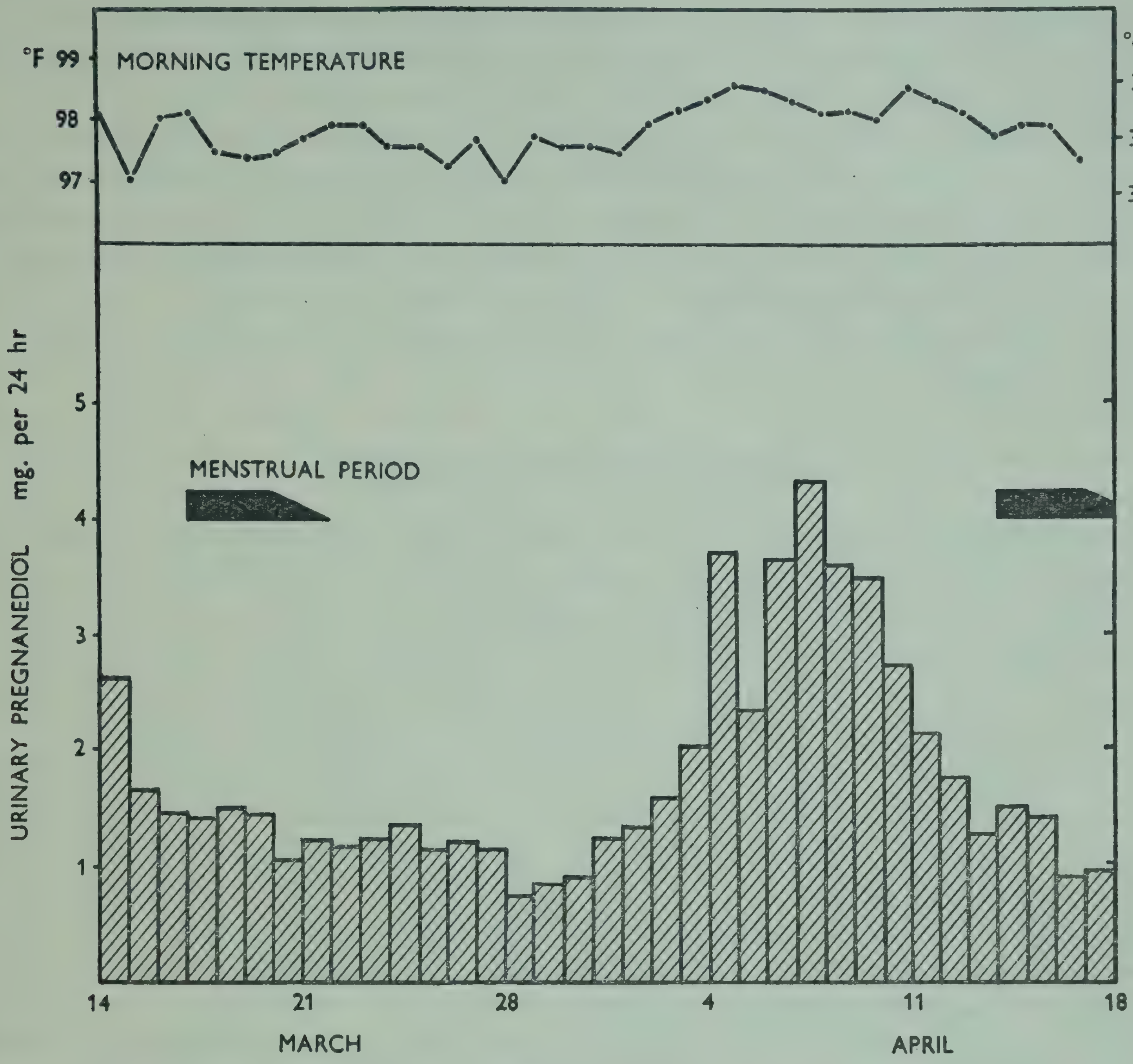


FIG. 52, 15

Pregnanediol excretion in the normal menstrual cycle.
(*By courtesy of A. Klopper.*)

an ovary containing ripe follicles but no corpora lutea. Although the lutein cells have for a long time been supposed to be the site of elaboration of progesterone this view, in the light of these recent findings, may have to be modified.

In the rabbit, Robson has shown that the corpus luteum can be maintained after hypophysectomy by suitable dosage of œstrogen. If this finding is applicable to woman the cessation of œstrogen production towards the end of the menstrual cycle is the cause of the degeneration of the corpus luteum and the drop in the excretion of pregnanediol.

The building up of the endometrium in the early part of the menstrual cycle must be attributed mainly to œstradiol and in the later stages to both œstradiol and progesterone. The decline in output of these two hormones at the end of the cycle, by withdrawing the stimulation, results in degeneration of the endometrium with menstrual bleeding.

The relationship of the pituitary hormones to ovarian activity is discussed below.

GONADOTROPHIC HORMONES

As a result of the work of Smith and Engle and also of Zondek and Aschheim, from 1927 onwards, a great deal has been learned about the controlling influence of the adenohypophysis on the ovary and testis. The hormones involved are termed *gonadotrophins.* If the pituitary gland is removed from an immature animal the gonads remain infantile and, since the development of the secondary sex organs, the clitoris, uterus and vagina in the female and the penis, prostate and seminal vesicles in the male, is dependent on the hormones of the gonads, these organs also remain immature. We have seen already that the secondary sex organs atrophy after gonadectomy (castration) in spite of the presence of the pituitary gland, and we can make the important generalization here that the adenohypophysis has no direct influence on the secondary sex organs but only an indirect one through the gonads.

The starting material for investigating the gonadotrophic hormones is usually pituitary glands obtained at the abattoir and dried by acetone. From both male and female glands two main gonadotrophic hormones have been extracted and prepared in apparently pure physicochemical form. One is called the *follicle-stimulating hormone* or FSH; the other is the *luteinizing hormone* or LH. Pure preparations of FSH, a glycoprotein, have been available only since 1949 and, as the effects of injection are modified by contamination with LH, the early accounts of the actions of impure extracts have to be discarded. Furthermore, since effects of pure FSH on an intact animal are complicated by the animal's own LH, it is necessary to study the effect of injection into hypophysectomized animals. Pure FSH prepared from the pituitary of the sheep or the hog given to hypophysectomized rats causes only slight growth of ovarian follicles with only a small output of œstrogens and no repair of the interstitial cells. Full œstrus is not obtained unless chorionic gonadotrophin is given at the same time. In hypophysectomized male rats FSH can maintain the weight of the testis and the production of motile spermatozoa but the output of androgens, as judged by the size of the accessory organs, is small.

LH is a glycoprotein varying in molecular weight and in composition according to the species from which it is obtained (page 1003); it has been available in a highly purified form since 1939. In the hypophysectomized female rat it can repair degenerated interstitial cells of the ovary and increase the output of œstrogens but it does not stimulate growth of follicles. In the male hypophysectomized rat LH stimulates the interstitial cells of the testis and increases the output of androgens as shown by the increase in size of the seminal vesicles and prostate; for this reason the hormone is often called the *interstitial cell stimulating hormone* or ICSH.

The separate actions of FSH and LH are only of theoretical interest since they always act together and their synergic action can be shown if the two pure hormones are injected at the same time. The train of events is probably as follows. FSH initiates growth of an ovarian follicle which matures under the influence of LH; at the same time the œstrogen output increases. Ovulation is induced and a corpus luteum is formed.

Up to the age of 12 years the girl excretes less than 1 rat unit of gonadotrophin per day but at puberty the daily excretion increases to 15 units and remains at this level till the menopause unless pregnancy supervenes when there may be a ten thousand fold increase (p. 980). After the menopause the daily excretion rises to about 75 rat units and then gradually declines. The gonadotrophic material of menopausal urine resembles FSH in that luteinization does not follow its injection into animals. It is believed, therefore, that the female menopause is due mainly to an unresponsiveness of the ovaries to pituitary gonadotrophin and not to a lack of these hormones.

In pregnancy very large amounts of gonadotrophic material are excreted in the urine but this differs in certain respects from pituitary gonadotrophic extracts. For example, pituitary extracts produce quite marked ovarian growth whereas those derived from pregnancy urine usually have mainly a luteinizing action. For this and other reasons it is believed that the source of gonadotrophin in pregnancy is not the pituitary but the placenta, and it is usually referred to as *chorionic gonadotrophin* or CG. Cell-free filtrates of human placenta grown in vitro possess gonadotrophic activity and it is likely that in man the cytotrophoblast (Fig. 52, 18) is the source of CG. The active principle appears to be a glycoprotein of molecular weight 100,000. Although preparations with an activity of 12,000 i.u. per mg. have been made no pure substance is yet available. CG given to hypophysectomized rats leads to stimulation of the interstitial cells but no follicular development. When CG is given to an intact rat it produces, by synergism with the animal's own pituitary gonadotrophins, ovarian hypertrophy with growth of the follicles and formation of corpora lutea as seen in the pregnancy test (p. 978). In male rats injection of CG causes hypertrophy of all the testicular elements. The effect of CG in man is very slight. The mechanism for the control of CG formation by the placenta is not known. It is only in woman that large quantities of gonadotrophin are excreted throughout pregnancy; in the chimpanzee and monkey and giraffe CG is excreted in the urine in early pregnancy only.

The serum of pregnant mares from the 45th to the 150th day of pregnancy is a rich source of gonadotrophin (PMS) which is produced by the endometrium; PMS is not excreted by the kidney and, therefore, remains in the blood for a considerable time. Injected into hypophysectomized or intact animals it produces an increase in the size of the ovaries with growth of follicles and sometimes formation of corpora lutea; PMS also stimulates the interstitial cells of the testis in animals. The human ovaries do not react so readily to PMS and its use in medicine is restricted by its antigenic properties (p. 378); in fact it has yet to be demonstrated conclusively that gonadotrophins can cause ovulation in women. The active principle of pregnant mare serum is a glycoprotein but it has not yet been prepared in a pure form.

Variations in sexual activity are associated with histological changes in the adenohypophysis. There is, for example, a degranulation of both acidophil and basophil cells after ovulation in the rabbit and at œstrus in several other animals. After castration in man and other animals the basophils become relatively more numerous; in certain animals some of these cells become enlarged and clear and are called *castration cells*. These changes in the basophils can be prevented by giving œstrogens. It has been suggested that the basophil cells are the main site of production of FSH, and that ICSH is secreted by acidophilic cells (see Fig. 53, 5).

The anterior pituitary seems to be under nervous and hormonal control. In the unanæsthetized rabbit electrical stimulation of the tuber cinereum (but not of the pars distalis) for a few minutes causes ovulation provided that the pituitary stalk is intact. If the stalk is completely severed neither mating nor electrical stimulation can produce ovulation and the gonads atrophy. In rodents after transection of the stalk œstrous cycles continue provided that the hypophysial portal vessels regenerate (see p. 1001) but not if they are prevented from crossing the gap by inserting a disc of paper in the cut. Bilateral lesions in the anterior part of the hypothalamus produce in female rats and guinea-pigs a continuous œstrus and in male rats failure of spermatogenesis and testicular atrophy. However, tumours responsible for precocious puberty in children are usually situated posteriorly in the mamillary bodies. Harris suggests that the hypothalamus regulates the activity of the adenohypophysis by hormones passing into the hypophysial portal vessels which arise in the median eminence (p. 1001).

In many mammals there is a seasonal variation in reproductive activity which seems to be dependent on environmental conditions such as light and temperature which excite or inhibit the anterior pituitary. For example, if in the course of the winter, when they are normally quiescent, female ferrets are given experimentally a summer ratio of hours of daylight to hours of darkness they come into heat in a few weeks. The mare is in the same category but the sheep and the goat respond to a decrease in the ratio of daylight to darkness. The pathway of the nerve impulses from the eye to the hypothalamus which activates the anterior pituitary is not known but the visual cortex is not involved.

Since after hypophysectomy the gonads atrophy there is no doubt about the effect of the pituitary hormones on the testes and ovary. There is a considerable amount of evidence also for a reciprocal action of the gonads on the pituitary; for example, injection of either œstrogenic or androgenic hormones causes, either directly or indirectly through the hypothalamus, an inhibition of pituitary activity which is shown by a fall in the gonadotrophic content of the pituitary and a reduction in the urinary excretion of gonadotrophins. This interaction between the gonads and the pituitary may be the basis of the regularly recurring œstrous cycles in rodents. The cycle may be considered as beginning with an active pituitary phase in which the gonadotrophic hormones released from the pituitary reach the ovaries and stimulate them to produce œstrogen in increasing amounts. After a time the œstrogen level in the blood rises sufficiently to cause inhibition of the pituitary, so that there is a drop in the output of FSH from the pituitary (and an increase in output of LH), and this in turn leads to a fall in œstrogen output. When the œstrogen level in the blood falls the pituitary becomes active once more and a new cycle begins. This explanation, however, is not the whole story. Pituitary glands transferred to the temporal lobe of the brain acquire a good blood supply but they do not support œstrous cycles. Further, ovaries transplanted into mature castrated male rats do not show cyclical changes but pituitaries from adult male rats implanted under the median eminence of hypophysectomized female rats support œstrous cycles and permit pregnancy to occur. These findings can be most easily explained by supposing that the anterior pituitary is entirely under the control of the hypothalamus and that the pattern of hormonal secretion depends on hormones reaching it via the hypophysial portal vessels.

HORMONAL BASIS OF PUBERTY AND THE CLIMACTERIC

An immature female rat can be brought to sexual maturity by implantation of tissue from the adenohypophysis of mature animals. After this procedure the animal goes into heat and allows mating. This experiment suggests that puberty is brought about by the discharge of gonadotrophic material from the adenohypophysis. In the immature male rat pituitary transplants can produce development of the seminiferous tubules but not spermatogenesis. The explanation of the delay in the occurrence of puberty cannot be that the pituitary does not produce gonadotrophins till that time because it has been found that the adenohypophysis contains a small amount of gonadotrophin even at birth, and implantation of pituitaries of immature animals into immature female rats can also bring about the changes of puberty. Furthermore, it is now known that if pituitary glands taken from rats a few days after birth are implanted below the pituitary stalk of hypophysectomized adult female rats, they become vascularized by the hypophysial portal vessels and full adult reproductive function is maintained. The most likely explanation for the delay in the appearance of sexual maturity is that, although the immature adenohypophysis contains gonadotrophins, these are not released from it in sufficient quantity until puberty to bring about maturation of the gonads. Even so, the stimulus responsible for the increased secretory activity of the anterior lobe of the pituitary at the time of puberty has still to be found but a clue may be provided by the observation that tumours responsible for precocious puberty most often involve the mamillary bodies. Poor nutrition delays the onset of puberty probably by depressing pituitary function (cf. p. 954.)

The gonads themselves play a part in determining the onset of puberty since it is only when the animal has arrived at this stage of its growth that the gonads begin to respond readily to gonadotrophic hormones. That there is a parallel increase in sensitivity of the secondary sex organs as puberty approaches is indicated by the fact that in immature animals large doses of androgens have little effect on the seminal vesicles, while at puberty quite small doses can cause an increase in size; similarly very large doses of androgens are necessary to induce spermatogenesis in immature animals but, once the process has begun, it can be maintained in mature hypophysectomized animals with a lower dosage. All these observations suggest that an increased production of pituitary hormones together with an enhanced responsiveness of the gonads and secondary sex organs results in the changes of puberty.

The phenomena of the climacteric in women are not due simply to decline of œstrogen production. There is at this time a temporary rise in the excretion of œstrogens which is followed by a gradual decline but even older women still excrete some œstrogenic material of unknown origin. After the menopause the gonadotrophin content of the pituitary is increased and the urinary output of gonadotrophins rises (p. 969); this increased output is presumably related to the falling off of ovarian activity, since it can be reduced to the premenopausal level by administration of natural or artificial œstrogens. After the menopause there is usually a fall in basal metabolic rate (see Table 11, 16) and a rise in body weight. Both changes could be attributed to a fall in thyroid activity.

In men at the end of reproductive life there is similarly an increase in the gonadotrophic substance in the urine but it is noticeable only after seventy years of age.

Thus the changes at the end of reproductive life seem to depend partly on a decreased power of the gonads to produce their hormones and partly on a diminished ability of the gonads to respond to the pituitary hormones.

OVULATION AND FERTILITY

The determination of the time of ovulation in woman is a very interesting physiological problem and one which has important implications in the study and treatment of sterility. An ovum can be fertilized at a mating which precedes ovulation only if the spermatozoa placed in the genital tract survive until ovulation occurs; fertilization can occur after ovulation only so long as the ovum remains capable of being fertilized. If, as is almost certainly true, the period of viability of both spermatozoa and ova is short the time of ovulation determines to a large extent the time during which mating is fertile.

Viability of Spermatozoa and Ova.—Microscopic examination of spermatozoa gives little information as to their fertilizing power; at most it can be said that semen containing a large number of abnormally shaped spermatozoa is unlikely to bring about a pregnancy. Normal spermatozoa retain their motility long after they have lost the power to

---

Fig. 52, 16

Decidua (the modified endometrium) at the 10th week of pregnancy. Stained with hæmatoxylin and eosin.

A. An area at a low magnification (× 13) showing the thickened uterine lining divided into a deep *stratu* *spongiosum* and a superficial *stratum compactum*.
B. An area in the *stratum compactum* showing the greatly enlarged stromal or *decidual* cells (× 70).
C. An area in the *stratum spongiosum* showing the large *pregnancy glands* (× 70).

(*By courtesy of W. W. Park.*)

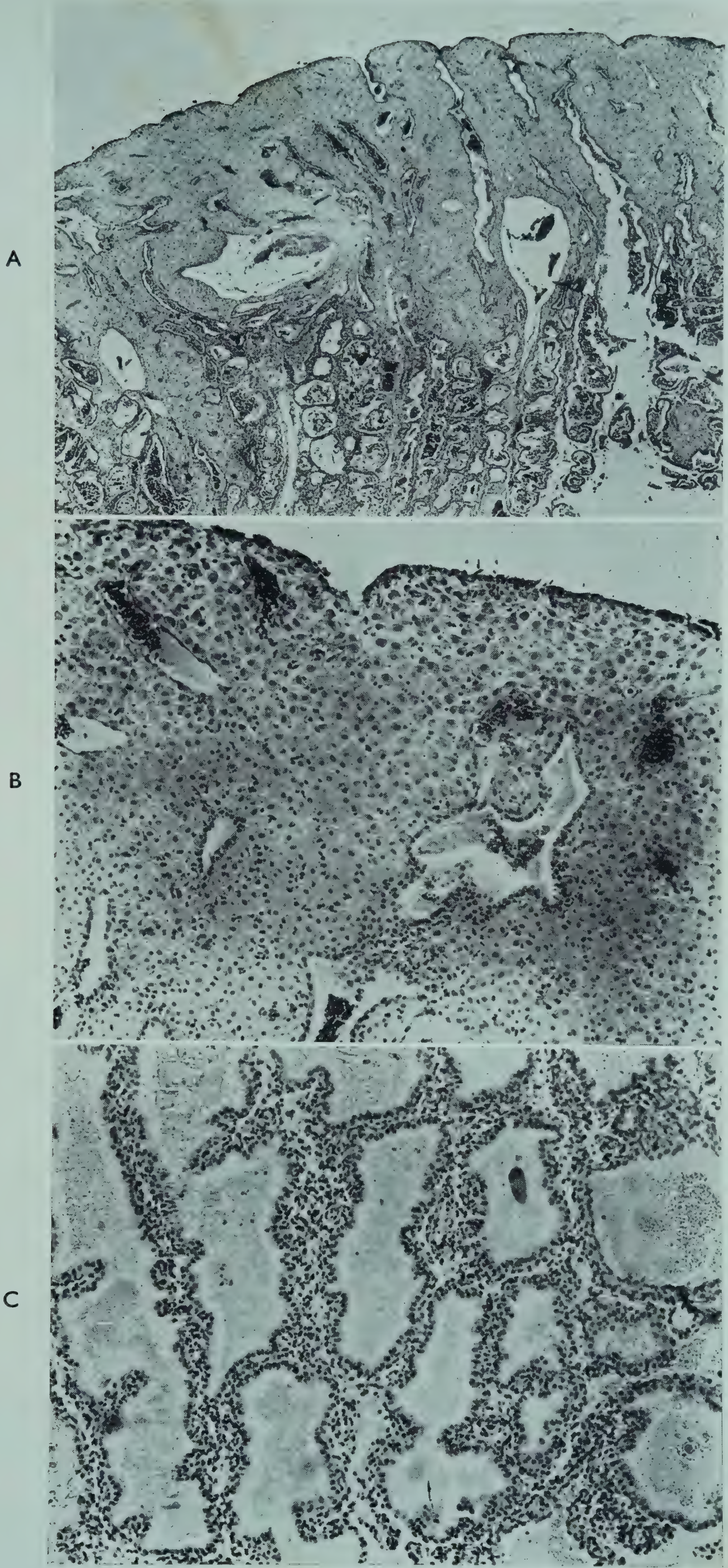


FIG. 52, 16

fertilize; direct experiments are therefore required in order to solve the problem of viability and these have been made in the rabbit by Hammond. The doe (female rabbit) remains in a continuous state of œstrus during the summer. In this species ovulation does not occur spontaneously but only some ten to twelve hours after mating. If a doe is mated with a buck rendered sterile by ligation of the vasa deferentia she ovulates and becomes pseudopregnant. During the sixteen days for which pseudopregnancy lasts she will mate, although mating is no longer followed by ovulation. If towards the end of pseudopregnancy she is mated with a normal buck (mating A) spermatozoa are deposited in the genital tract but no ovulation occurs. When pseudopregnancy is ended mating with a sterile buck (mating B) produces ovulation without insemination. Hammond found that if the interval between A (deposition of spermatozoa) and B (ovulation) was greater than four days none of the does became pregnant, while pregnancy occurred in some animals if the time between A and B was slightly less than four days. He concluded that the *maximum* life of a rabbit spermatozoon is four days. One factor limiting the survival of spermatozoa after insemination is the relatively high temperature of the vagina and uterus as compared with that within the scrotum.

The ovum has an even shorter life than the spermatozoon. Histological examination shows that the unfertilized ovum quickly degenerates in the uterine tubes. Hammond mated doe rabbits with sterile bucks to produce ovulation without insemination; subsequent matings with fertile bucks later than six hours after ovulation produced no pregnancies, showing that the maximum life of the rabbit ovum is about six hours. Experiments in which guinea-pigs were inseminated at various intervals after ovulation give a maximum life of thirty-two hours for the ovum although eight hours should probably be regarded as its usual period of viability in this animal. There is reason to believe that the primate ovum remains capable of fertilization for a short time only. Data obtained from animal experiments must not be applied without qualification to human beings but they suggest that if a mating is to be fertile it must occur within a day or so of ovulation.

TIME OF OVULATION.—In a number of mammals, including the rabbit and cat and possibly the rat, coitus produces reflexly a release of LH from the adenohypophysis and so brings about ovulation but it is almost certain that ovulation in woman occurs spontaneously, that is, is not induced by coitus. The finding of an actual ovum is, of course, the best proof that ovulation has recently occurred and in a few cases human ova have been obtained by washing out the uterine tubes and collecting the fluid which emerged from the fimbriated end; the ova were found only at the middle of the cycle (about the fifteenth day) and not at other times. Inspection of human ovaries during abdominal operations has shown that ovulation usually occurs between the twelfth and fifteenth day of the menstrual cycle. The change from the proliferative to the secretory phase of the endometrium which depends on the presence of a corpus luteum also occurs about the middle of the cycle. Although it is difficult to obtain information about mating in the human subject analysis of cases in which conception resulted from a single insemination on a known day of the cycle shows, on the assumption that the ovum is short-lived, that ovulation can occur at any time between the sixth and twentieth day of the cycle. In agreement with this is the fact that artificial insemination in women is usually most successful about the middle of the cycle, although some women become pregnant only when inseminated during the first week. It has been claimed that the slight fall of body temperature followed by a slight rise, usually less than 0·5° C, which occurs about the middle of the menstrual cycle (Fig. 52, 17) indicates the time of ovulation but, as a result of direct tests of the validity of this claim by inspection of the ovaries, it is now known that the rise of temperature may occur before ovulation or may be delayed as much as three days after ovulation. The slight elevation of body temperature in the second half of the cycle has been attributed to the presence of progesterone in the circulation; at any rate it can be imitated by giving progesterone after a preparatory course of œstrogens in women with amenorrhœa (absence of menstruation).

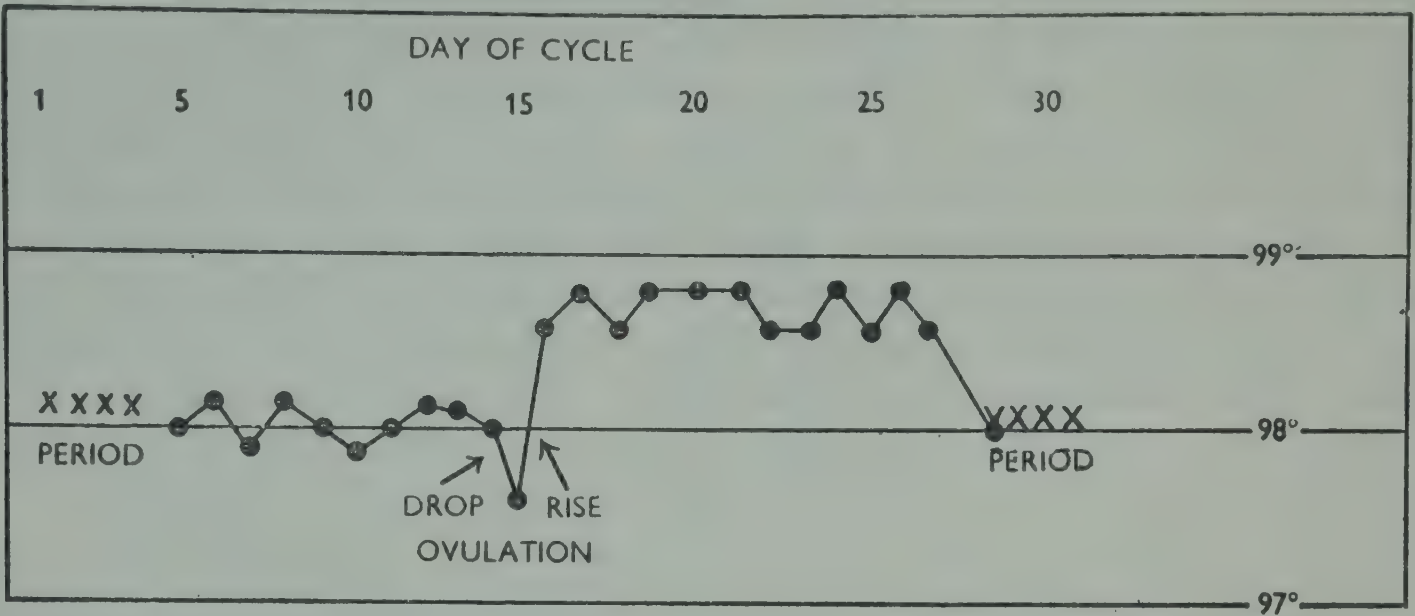


FIG. 52, 17

An idealized basal temperature chart in which the temperatures were taken by the patient before she rose in the morning or took foods or liquids. Crosses indicate menstruation (P. Tompkins (1945), *J. Obstet. Gynœc. Brit. Emp.* **52,** 243).

PREGNANCY

Conception occurs after a fertile mating in which the spermatozoa deposited in the vagina reach an ovum shortly after ovulation. Before coitus the penis becomes erect due to the engorgement of its blood vessels brought about chiefly by the dilatation of the arterioles but partly by the contraction of the ischiocavernosus muscle which obstructs the venous outflow from the organ. In the woman the clitoris also becomes erect and the vulva and vagina secrete mucus which facilitates the passage of the penis into the vagina. Coitus culminates in the orgasm during which the semen is ejaculated by rhythmic contractions of the bulbocavernosus and ischiocavernosus muscles. During orgasm in the female, which should occur simultaneously, it is likely that the os uteri dilates, because spermatozoa have been found in the uterus a few minutes after coitus. Coitus is accompanied by psychic excitement and must be regarded primarily as a cerebral event although the reflex changes in the genitalia are governed by lumbar and sacral spinal centres. Stimulation of the sacral (parasympathetic) outflow in the pelvic splanchnic nerves (p. 804) causes erection of the penis or clitoris by dilatation of their blood vessels and these nerves have, therefore, been called the *nervi erigentes.* Somatic nerves arising in the sacral region of the cord pass out in the pudendal nerves to the ischiocavernosus and bulbocavernosus muscles in both sexes and, therefore, both somatic and autonomic nerves are concerned in producing erection.

The internal organs of reproduction in both sexes are supplied by sympathetic fibres from the lumbar outflow travelling in the hypogastric nerves. Stimulation of these nerves in nonpregnant women produces a contraction followed by an inhibition of the body and cervix of the uterus. After section of the hypogastric nerves in man the power of ejaculation but not of erection, is lost.

Although only one spermatozoon is necessary for the fertilization of an ovum, pregnancy does not occur unless many millions are deposited in the vagina. It has been suggested that it is necessary to have such a large number to supply sufficient hyaluronidase (p. 17) to decompose the jelly-like hyaluronic acid which glues together the cumulus cells surrounding the ovum, but experiments by Shettles in 1953 on human ova obtained from ovarian follicles or by washing out the uterine tube seem to disprove this theory. Incubation of such ova with semen or with hyaluronidase had little action on the cumulus cells but incubation with tubal mucosa caused denudation of the ovum. But the cumulus cells may not be such a barrier as has been supposed for the phase contrast microscope has shown male and female pronuclei and sperm tails in newly fertilized rat ova which were still covered with cumulus cells. Incidentally Shettles found no evidence that the ova possessed a trophic influence on spermatozoa. The latter seem to move about quite randomly and sometimes passed by quite close to the ovum. If however a spermatozoon touched the ovum it remained in contact. Fertilization occurs usually in the uterine tubes, probably within 12 hours of coitus in man, and the ovum then passes either by peristaltic or by ciliary action into the body of the uterus.

In the cow spermatozoa reach the uterine tubes two and a half minutes after insemination—during this time they must travel about 50 cm. Since this far exceeds their unaided rate of progress the spermatozoa must be transported by uterine contractions which may be excited by oxytocin released from the neurohypophysis. The observations that coitus in man and animals causes ejection of milk from the mammary glands and that in cows mating causes strong uterine contractions afford circumstantial evidence for this release. It is known, however, that conception may occur after removal or denervation of the neurohypophysis.

When an ovum is fertilized the corpus luteum persists instead of degenerating at the end of the menstrual cycle and menstruation does not occur. The hormonal or nervous mechanism which is responsible for this is not known with certainty but it is likely that chorionic gonadotrophin (p. 969) is produced very early by the embedding blastocyst in sufficient quantity to maintain the corpus luteum. Failure to menstruate at the expected date is thus one of the first presumptive signs of pregnancy in woman. The developing ovum at the blastocyst stage reaches the uterus in four days or so after ovulation when the endometrium is in the premenstrual phase and sinks into it about the seventh day by digesting away the superficial layers; this process may also provide food materials for the growth of the blastocyst. The endometrium continues to grow (Fig. 52, 6) during pregnancy, since there is no menstrual degeneration, and it may eventually reach 10 mm or more in thickness. The three layers described in the premenstrual endometrium become more obvious (Fig. 52, 16). The superficial layer of the stroma becomes compact and the greatly enlarged stromal cells are known as *decidual cells.* In the much thicker spongy layer below this the glands are widely dilated so that there is little stromal tissue. In the limiting or boundary layer next the uterine muscle the glands are tortuous but are not dilated. Since a large part of the endometrium is cast off with the placenta at parturition it is often referred to as the *decidua*; after parturition the endometrium is re-formed from the limiting layer. In the decidua basalis under the developing embryo there is from an early stage a great dilatation of the maternal blood vessels and small finger-like outgrowths of the outer layer of the blastocyst, the chorionic villi, grow into them by erosion of the decidua. This penetration of the villi is aided by obliteration of small arteries of the decidua causing necrosis and the formation of large spaces in the decidua which fill with maternal blood. As the villi are soon invaded by mesoderm carrying fœtal blood vessels the fœtal and maternal circulations are brought very close to one another and in this way the placenta is formed. The placenta in man is described as hæmochorial, since the chorionic villi dip directly into maternal blood spaces. There is, however, no direct connexion between the two vascular systems and blood cannot pass from mother to fœtus or vice versa. Intervening between the fœtal and maternal circulations are only three thin layers (Fig. 52, 18), namely the fœtal vascular endothelium, the connective tissue of the villi and the trophoblast, across which the transfer of nutrient materials readily occurs.

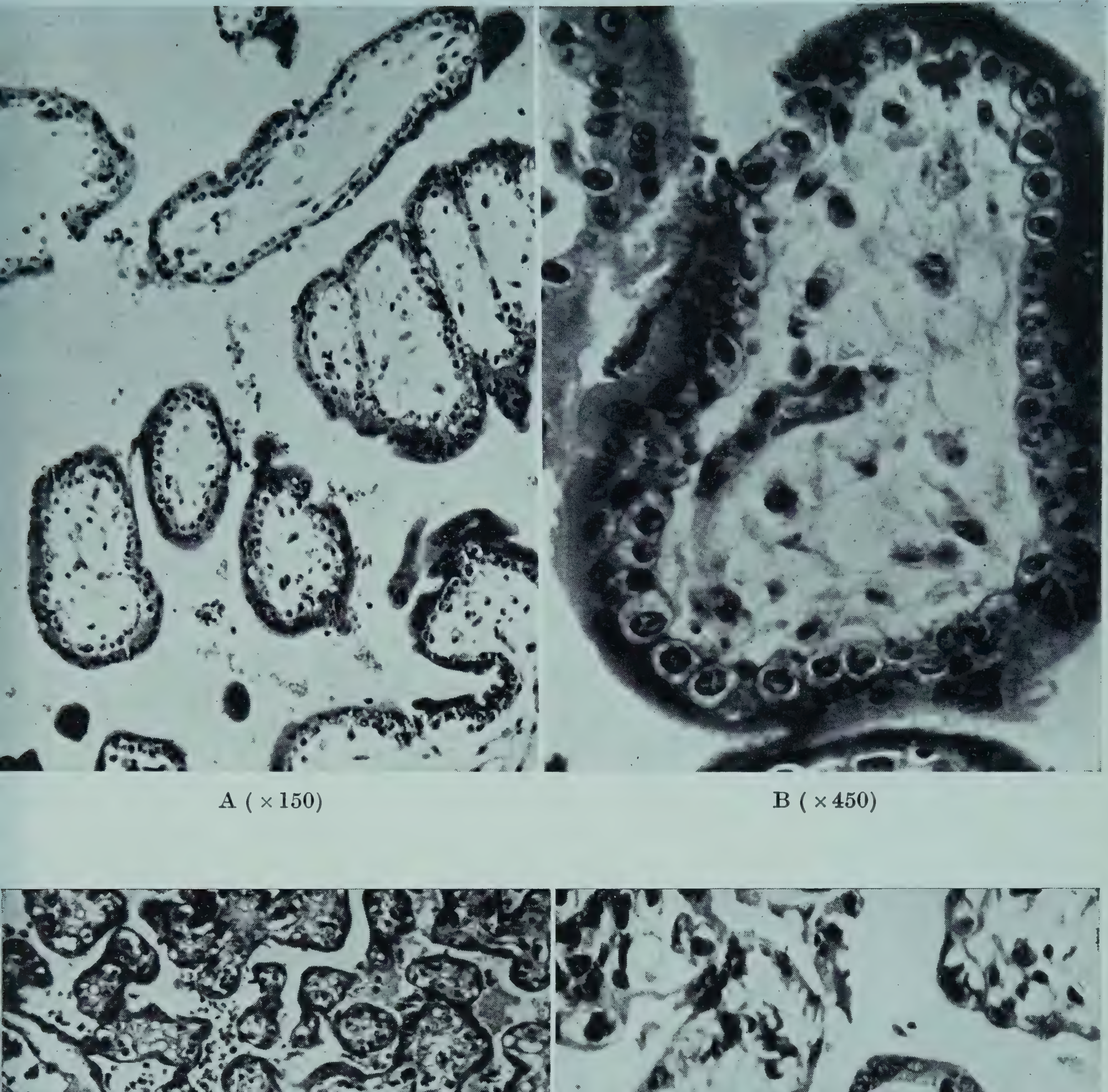
As pregnancy advances the products of conception increase in size and eventually occupy the whole of the uterine cavity. The growing fœtus attached by its umbilical cord to the placenta is bathed by the amniotic fluid which is contained within the 'bag of membranes' consisting of the amnion and chorion. Amniotic fluid is fairly freely exchanged with the maternal fluids; it is mainly an excretory product of the amniotic epithelium but it also contains urea and bile salts derived from the fœtus. The function of the amniotic fluid is to provide space for fœtal growth and movements and to distribute the pressure due to uterine contractions evenly over the fœtus.

FIG. 52, 18

Sections of human placental tissue.

A and B show chorionic villi at the 10th week of pregnancy.
C and D show chorionic villi at term. The villi have more numerous and finer terminal branchings than in the earlier placental tissue. Hæmatoxylin and eosin.
The early villus consists of a core of mesodermal tissue containing an occasional blood vessel surrounded by an inner and an outer layer of trophoblastic epithelium. The cells of the inner layer are relatively large and clear and form the Langhans' cell layer or *cytotrophoblast.* The outer layer is composed of smaller, darker nuclei lying irregularly in a background of cytoplasm which shows no cell boundaries: this is the syncytium or syncytial layer or *syncytiotrophoblast.* As the villus ages the cells of the cytotrophoblast become fewer and usually disappear entirely, leaving only the irregular syncytiotrophoblast as a covering for the villus: at the same time the blood vessels within the core of the villus gradually increase in number until they may occupy 50 per cent. of the cross-sectional area of a villus during the last weeks of pregnancy.

(*By courtesy of W. W. Park.*)


A (× 150)

B (× 450)

C (× 150)

D (× 450)

Fig. 52. 18

[*Facing page* 976

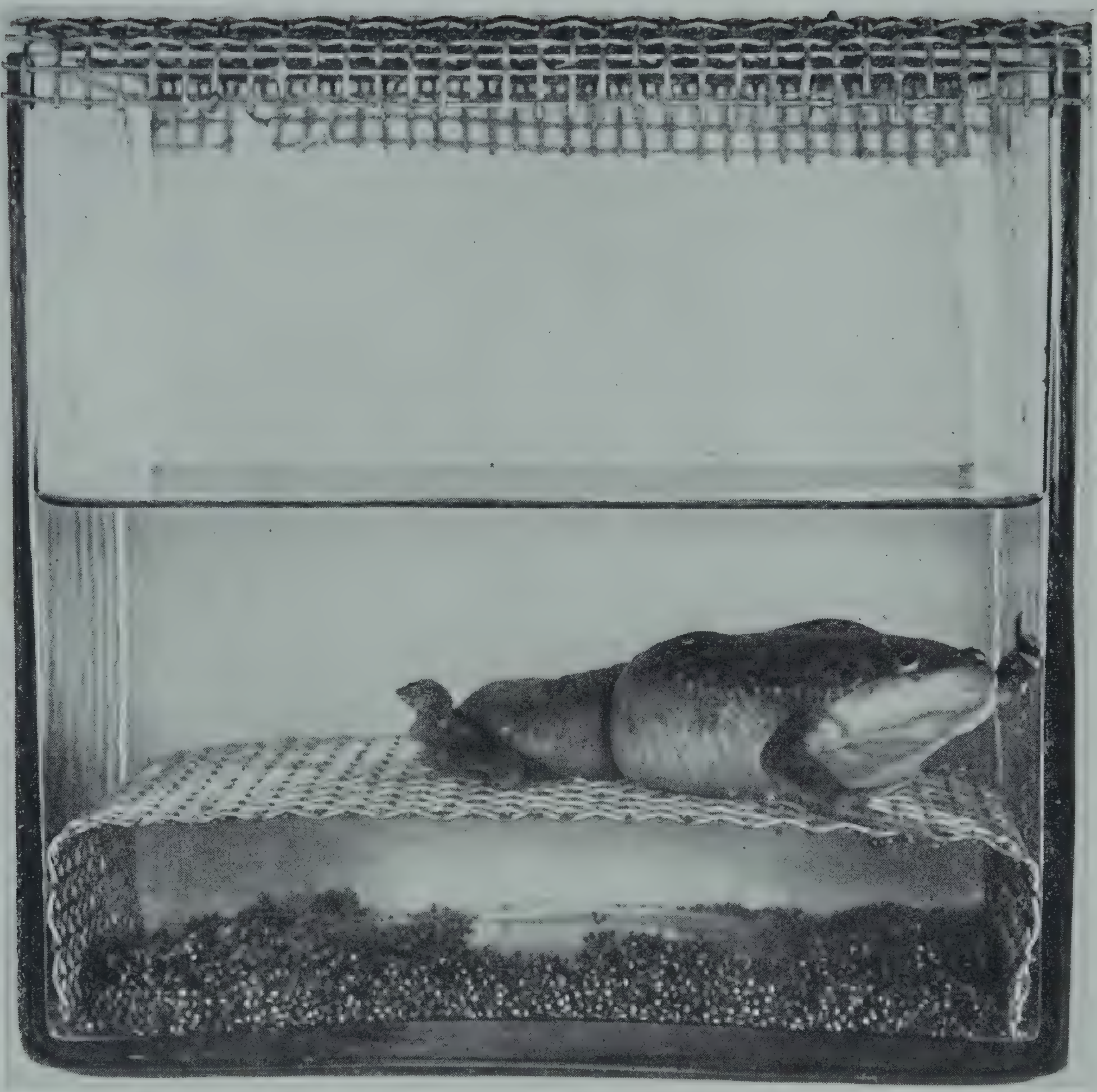


FIG. 52, 19

The toad *Xenopus* after injection of pregnancy urine. The extruded eggs pass through the wire netting and accumulate on the floor of the glass jar.

(*By courtesy of J. M. Dodd.*)

The Hormonal Control of Pregnancy

In some animals, such as the rabbit, cat and dog, hypophysectomy at any stage of gestation brings the pregnancy to an end. In others, such as the mouse, rat, guinea-pig and monkey, the pituitary can be removed at least in the later stages of pregnancy without producing abortion; indeed in the mouse and monkey the mammary glands develop normally and delivery occurs at the expected time, pituitary deficiency being shown only by the absence of lactation. It is not known to which category the human subject belongs. In all species the ovary is necessary in the early stages of pregnancy since the hormone from the corpus luteum (progesterone) is responsible for preparing the endometrium for the reception of a fertilized ovum but after this preliminary stage the ovaries can be removed in several species, including man, monkey, cat and dog, without disturbing the pregnancy. In these animals it is likely that the placenta provides all the hormonal material necessary for the continuation of pregnancy. Since both œstrogen and pregnanediol are excreted by pregnant öophorectomized women there is presumptive evidence that œstrogen and progesterone are produced in the placenta as well as in the ovaries, and it is likely that the main site of production of these hormones during pregnancy is the placenta. Since the suprarenal gland contains both œstrogen and progesterone it may play a part in the maintenance of pregnancy. There is evidence in some animals at least that the placentæ maintain, possibly by producing œstrogens, the activity of the corpora lutea. Intraocular transplants of human placenta which show syncytiotrophoblast (see legend to Fig. 52, 18) produce œstrogenic stimulation of the uterus of the oöphorectomized rabbit. Progesterone has recently been isolated from the human placenta. The human placenta has a high gonadotrophin content, especially early in pregnancy, but the function of the gonadotrophic hormones is not clear; since a high excretion of these hormones occurs early in pregnancy in women (Fig. 52, 22), at a time when the corpus luteum is active, they may maintain the secretion of progesterone by the corpus luteum until this function is assumed by the placenta. In agreement with this idea is the finding that injection of chorionic gonadotrophin into non-pregnant women increases the pregnanediol excretion and delays the onset of the next menstrual period. The corpus luteum may therefore be maintained either by gonadotrophins or by œstrogens.

Urinary Excretion of Hormones in Pregnancy

Œstrogens.—The œstrogens are excreted in the urine chiefly as glucuronides, the daily urinary output rising during the latter part of pregnancy with considerable daily fluctuations. There is usually a fall in output in the last two days before the onset of labour; after parturition the excretion falls in a few days to the non-pregnant level (Fig. 52, 20). However, the amount of œstrogen eliminated in the urine is not a reliable guide to the amount produced in the body, since the proportion destroyed in the body, mainly in the liver, is quite unknown (see p. 965).

Pregnanediol.—The excretion of pregnanediol during pregnancy is shown in Fig. 52, 21. Here again the amount excreted gives no satisfactory clue to the actual production, since only a fraction of injected progesterone is recovered as pregnanediol glucuronide in the urine (p. 964). There is no doubt, however, that progesterone is produced in much larger quantities in pregnancy than it is in the luteal phase of the menstrual cycle. Pregnanediol is not excreted in any other species.

Gonadotrophic Substances.—The presence of gonadotrophins in the urine of pregnant women was first demonstrated by Aschheim and Zondek in 1928, and the detection of these substances forms the basis of all the reliable human pregnancy diagnosis tests. Quantitative measurement (Fig. 52, 22) shows that the rise in gonadotrophin (CG) excretion begins quite suddenly about the time of the expected, but missed, menstrual period, rises to a peak a month or so later and then falls to a much lower but still easily detectable value, around which it fluctuates until the end of pregnancy. The ovum is becoming embedded in the decidua at the time of the initial rise in gonadotrophic excretion but, although it is unlikely to be a mere coincidence, the connexion between the two events is as yet unknown. Within a few days after parturition the excretion falls to very low values not detectable by the usual pregnancy diagnosis tests. After it is injected gonadotrophic material from the pituitary is partly destroyed in the body and partly excreted and, therefore, the amount eliminated in the urine may be much less than that actually produced in the pituitary and placenta.

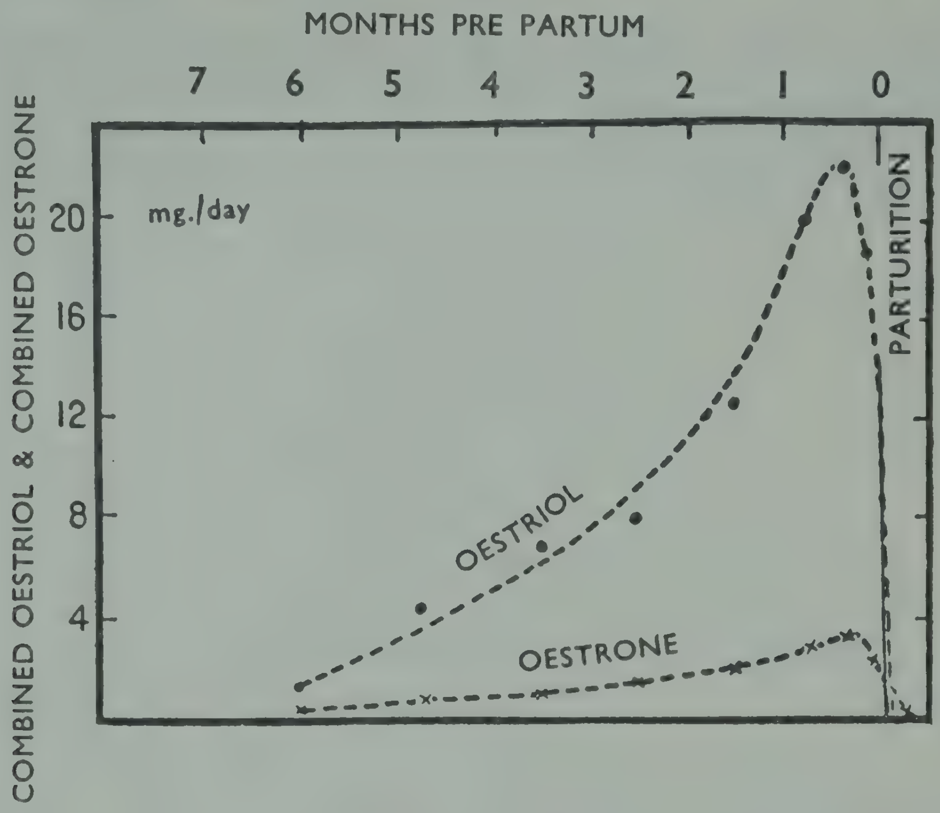


FIG. 52, 20

Excretion of œstrogen during pregnancy (average curves). All figures are mg./24 hours excretion (M. Cohen, G. F. Marrian & M. Watson (1935), *Lancet*, i, 675).

PREGNANCY DIAGNOSIS TESTS

In the original test devised by Aschheim and Zondek in 1928 urine from the pregnant woman, after removal of toxic material by extraction with ether, is injected into immature mice, which are killed five days later; the presence in the ovaries of corpora lutea or tiny hæmorrhages, or both, constitutes a positive diagnosis of pregnancy. The amount of gonadotrophin in the urine of non-pregnant women is not sufficient to bring about these changes in the ovaries. In the first weeks of pregnancy and occasionally in the late stages of a normal pregnancy when the urinary gonadotrophin content is relatively low there may be merely enlargement of the ovarian follicles, which appear as white spots, and an increase in the size of the uterus; such a result is not as a rule considered as positive.

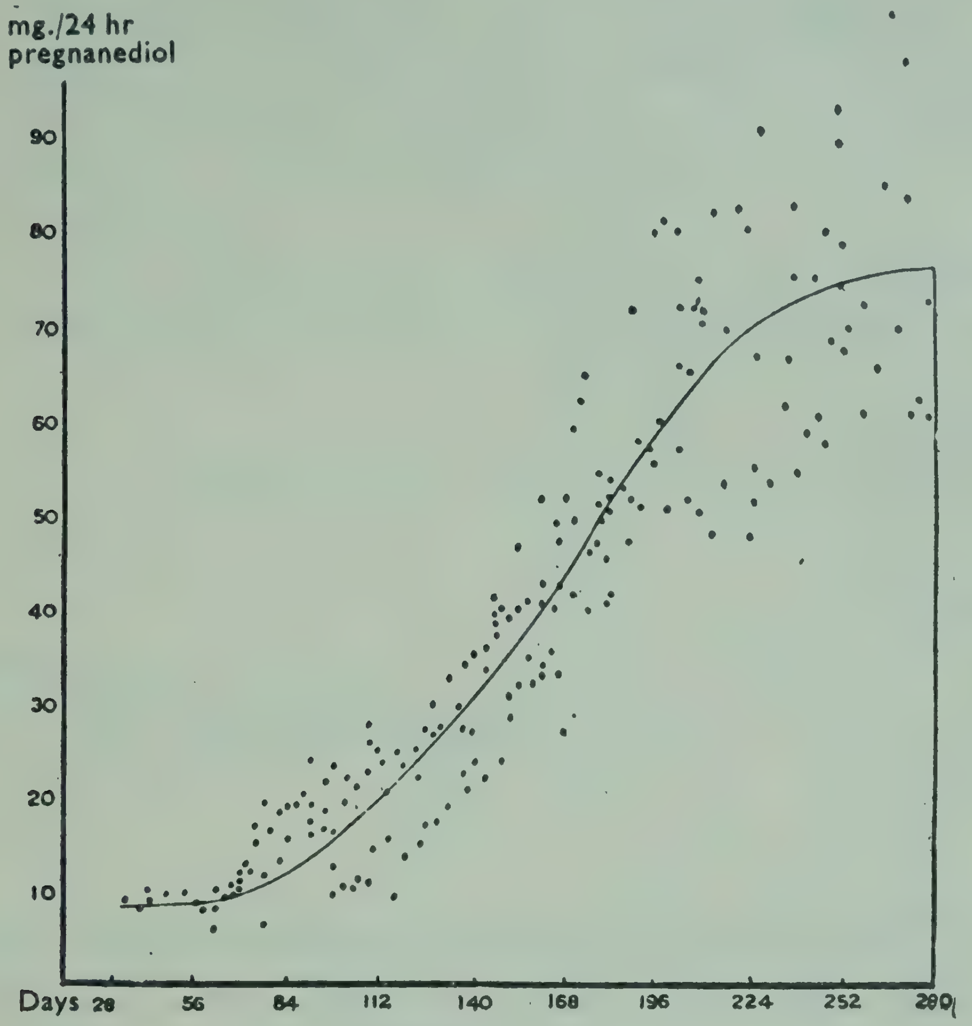


FIG. 52, 21

Daily excretion of pregnanediol in the urine of eight patients during normal pregnancy (J. S. L. Browne, J. S. Henry & E. H. Venning (1939), *Amer. J. Obstet. Gynecol.* **38**, 931).

For a biological test, the Aschheim-Zondek test is remarkably reliable, an incorrect result being returned in only one or two per cent. of cases.

Friedman showed in 1929 that the rabbit could be used instead of the mouse with the advantage that the test could be completed in 24 or at most 48 hours. Ten ml. of urine detoxicated by treatment with sulphosalicylic acid are injected into an ear vein of a rabbit which has been isolated to prevent mating. Hæmorrhagic or ovulated follicles constitute a positive diagnosis.

Hogben found in 1931 that the female South African clawed toad *Xenopus* can also be used to detect gonadotrophins. In this case the injection of an extract of pregnancy urine into a lymph sac produces extrusion of eggs usually within 12 hours (Fig. 52, 19). Male toads of many varieties can also be used to give a reliable and quick pregnancy diagnosis test, the presence of spermatozoa in the cloaca a few hours after the injection of urine constituting a positive diagnosis.

Other pregnancy tests depending on the assay of œstrogen, pregnanediol, or histidine in urine have been proposed but such tests are much less reliable than those described above.

There is a large excretion of CG in hydatidiform mole and chorionepithelioma, in both of which there is abnormal proliferation of chorionic tissue during or after a pregnancy. Occasional false positives are given in the menopause and diseases of the ovary and uterus in which there is an increase in the output of pituitary gonadotrophins. It should be noted that a positive pregnancy test is not a direct indication of a live fœtus but merely indicates that living cytotrophoblast is excreting CG into the blood.

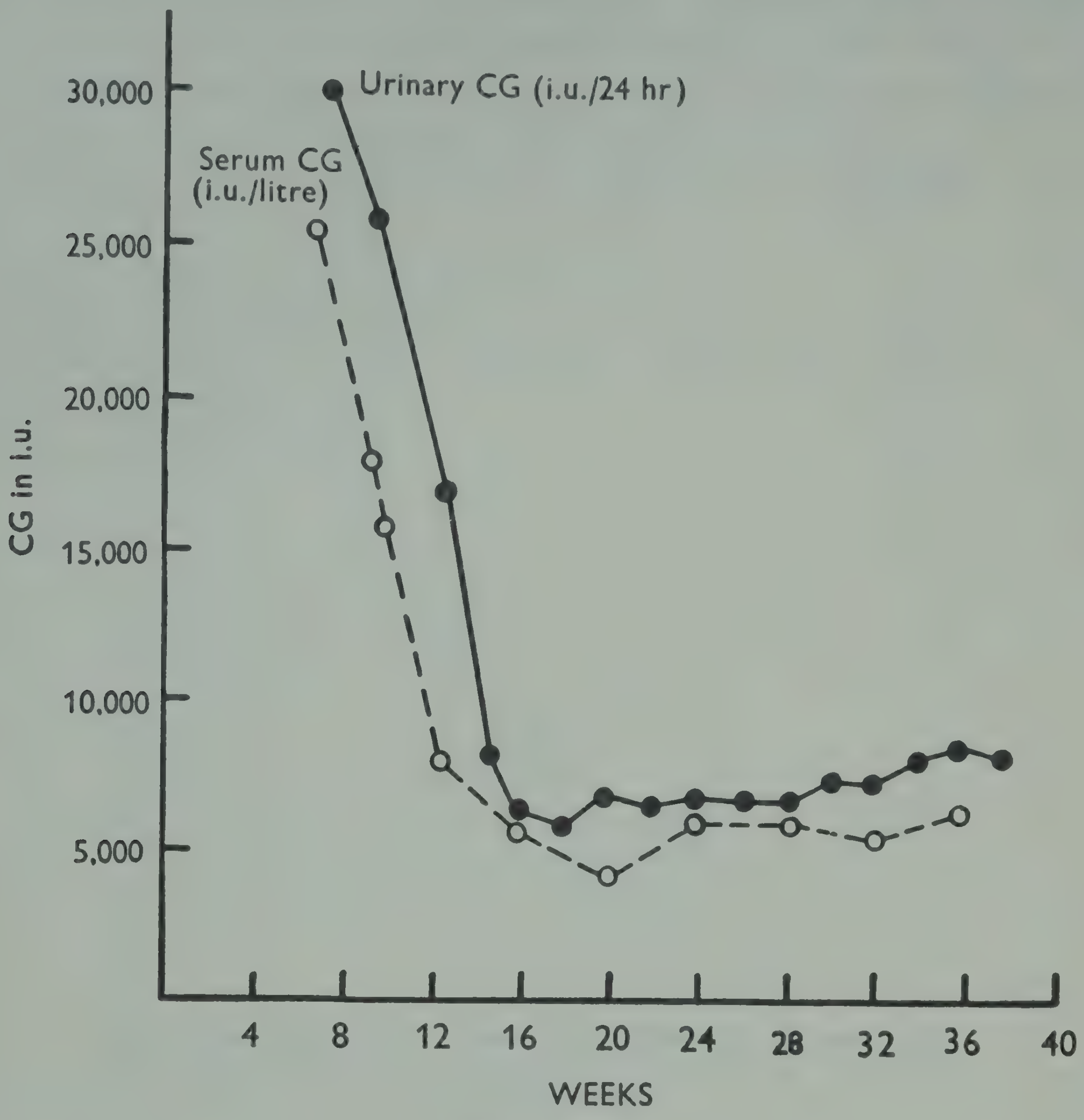


Fig. 52, 22

Mean values of urinary and serum chorionic gonadotrophin CG throughout normal pregnancy. CG was assayed by its effect on prostatic weight in rats. Total number of estimations in urine 136, in serum 140.

(*By courtesy of J. A. Loraine.*)

Water Retention in Pregnancy

Although it is difficult to obtain accurate figures for water retention in the complicated circumstances of pregnancy, the average gain in extracellular fluid has been assessed as about 6·3 litres distributed as follows (Chesley):

	litres
Products of conception	1·5 to 2
Uterus and breasts	0·7
Blood plasma	1·3
Extragenital interstitial fluid	2·5
	6·0 to 6·5

The diuresis after drinking water is smaller in a pregnant than a non-pregnant woman. During the greater part of pregnancy sodium is retained to the extent of 3 g. per week, and this retention would lead to a rise in the osmotic pressure of the tissues unless water were retained at the same time. The plasma volume is also considerably increased in pregnancy but, since the colloid osmotic pressure is actually some 20 per cent. lower than in the non-pregnant woman, it is difficult to see the reason for the increase. It is likely that this retention of sodium and water is related to the increased production of œstrogens in pregnancy since œstrogen therapy in women has produced an increase of up to 20 per cent. in blood volume with a lowered hæmatocrit.

The figures quoted above are for water only and do not give the total gain in weight during pregnancy which has been roughly estimated to be made up as follows:

	kg.
Products of conception (baby, 3·2; amniotic fluid, 0·7; placenta, 0·5)	4·4
Uterus	0·9
Breasts	0·7
Increase in blood volume	0·5
Gain in fluid and fat in maternal tissues	4·5
	11·0

This is equivalent to an addition of 22 per cent. to the body weight of an eight-stone (50 kg.) woman. An even greater average increase in weight, as much as 25 per cent. of the non-pregnant weight, has been reported by other investigators but many obstetricians now advise that the gain in weight during the whole of pregnancy should not exceed 10 kg. A sudden gain of weight, say 0·5 kg. or more in a week, is often a sign of impending toxæmia.

The gain in body weight during pregnancy has been investigated by Newton. Mice increase in weight during pregnancy and regain their initial weight after parturition. If the fœtuses are destroyed about the fourteenth day of pregnancy by applying pressure through the abdominal wall, the body weight is maintained at the high pregnancy level so long as the placentæ are retained. When the placentæ are expelled at full term (twentieth day) there is a rapid loss of weight and a return to the level before pregnancy; the loss in weight is probably due to a loss of body water and is about eight times greater than the weight of the expelled placentæ. Similar changes in body weight are observed in monkeys in which the fœtus has been removed without disturbing the placenta. The hormonal mechanism by which the placenta brings about water retention is not known.

FUNCTIONS OF THE PLACENTA

The early stages in the formation of the placenta have already been described (p. 976). The organ increases in size up to the thirty-second week of pregnancy and at full term weighs from one-quarter to one-twelfth as much as the child; a large placenta increases the chances of the survival of the child. Patches of degeneration commonly seen in the placenta at the end of pregnancy are apparently due to local coagulation of the maternal blood in the area affected. The circulation through the placenta is illustrated in Fig. 52, 24.

The endocrine functions of the placenta have been mentioned already (p. 977) and it is now necessary to consider the transfer of food materials and waste products between mother and fœtus. The supply of oxygen to the fœtus is clearly an important matter; the amount transferred depends first on the amount of oxygen brought to the maternal blood sinuses and on the volume and rate of the circulation in the chorionic villi. As pregnancy proceeds the uterus becomes much more vascular but in spite of the greatly increased blood supply to the uterus at the end of pregnancy the arteriovenous oxygen difference, in the rabbit at least, is much increased. The percentage oxygen saturation of the maternal blood leaving the uterus, that is mainly blood which has traversed the placenta, is about 70 in the first half of pregnancy but declines to about 20 just before parturition. In the anæsthetized goat at term the oxygen tension in the maternal blood going to the placenta is 60 mm Hg and that in the venous blood arriving from the fœtus is 15 mm Hg. The gradient of 45 mm Hg is sufficient to account for the diffusion of oxygen across the placenta and to make a secretory exchange most unlikely. The difference in carbon dioxide tensions, 61 mm Hg (in the fœtus) minus 43 mm Hg (in the mother), is more than adequate to account for exchange of this gas by diffusion. Furthermore when the carbon dioxide tension in the mother's blood is raised by asphyxia carbon dioxide may pass in the reverse direction, that is into the fœtus; this can only be explained on the basis of transfer by diffusion.

The maternal blood sugar is higher than the fœtal and a rise in blood sugar in the maternal circulation is followed by an increase in the fœtal blood sugar. Since the latter returns quickly to the resting value, the sugar must be utilized or stored in the fœtal tissues. Very little fructose is present in human fœtal blood but the fœtal sheep has both glucose and fructose in easily measured amounts in its blood. Glucose injected into a sheep fœtus still attached to the placenta is partly transformed into fructose; since this does not occur when the cord is tied the transformation must occur in the placenta. In a woman at the end of pregnancy, intravenous injection of glucose raises the fœtal fructose very little but injection of fructose raises both fœtal and maternal blood glucose. Glucose passes through the placenta from a zone of high concentration to a zone of lower concentration but the amount transferred is not proportional to the difference in concentration as would be expected in simple diffusion. A carrier system has been postulated but there is no clue as to its nature. The fœtal liver glycogen varies more rapidly than the glycogen stored chiefly in the decidual tissue of the placenta; the former fluctuates with the fœtal blood sugar while the latter is relatively constant in the absence of excessive tissue demands. It is believed that insulin does not pass across the human placenta but it does so in the rat at full term.

Fat and also a lipase have been found in the human placenta but there is no evidence that any hydrolysis of fat occurs there. By injecting phospholipids marked with ^{32}P into the circulation of pregnant rabbits it has been found that although the placenta takes up large amounts of phospholipids it does not transmit them to the fœtus. Fat seems to be able to pass through from the maternal to the fœtal circulation but the mechanism is not known. Some of the fœtal fat is derived from carbohydrate. When labelled inorganic phosphate ($Na_2H^{32}PO_4$) is injected into a pregnant rabbit, the plasma phospholipids are after a time labelled with ^{32}P. The radioactivity of the phospholipids extracted from the fœtal liver and placenta is much higher, four hours after injection, than that of the fœtal or maternal blood phospholipids. These two fœtal organs must therefore be sites of active synthesis of phospholipid.

The fœtal blood amino-acid concentration is always slightly higher than that of the mother but injection of amino-acids into the mother increases their concentration in the fœtal blood. The explanation of this movement across the placenta, apparently against the diffusion gradient, is not known. As might be expected from the rate of increase in fœtal weight, protein synthesis in the fœtus is at its maximum at the end of pregnancy and must depend on transfer of amino-acids across the placenta. Although there is a proteolytic enzyme in the placenta its function, if any, in protein transport is unknown. The passage of the Rh antigen from the fœtus to the mother and the return passage of the anti-Rh material into the fœtus (p. 382) show that unchanged protein must be able, although in nutritionally insignificant amounts, to pass between mother and child. Experiments by Brambell are against placental transfer and it may be that the exchange occurs through the non-placental chorion (chorion læve) and the amniotic fluid.

Even if the mother is anæmic the child is usually born with a normal hæmoglobin concentration. It may, however, quickly develop anæmia since even in the most favourable circumstances the fœtal stores of iron are relatively low. Radioactive iron studies have shown that the pregnant woman absorbs iron from the food much more rapidly than a non-pregnant woman and that the iron may pass directly to the fœtus. The fœtal blood calcium and inorganic phosphorus are always above the maternal level and the considerable demands made on the mother for calcium and phosphorus may exceed her intake of these materials; calcium and phosphorus are then drawn from the mother's bones and teeth. For these reasons the daily intake of calcium and phosphorus in the last three months of pregnancy should be at least 1·5 g. and 2 g. respectively.

Vitamin A passes through the placenta to the fœtal liver, especially if the diet contains fat. In deficiency of this vitamin in man and animals fertilization and implantation may fail and in some animals, if the deficiency is gross, fœtal malformation may occur. Both vitamin D and vitamin E also pass through the placenta. The former is stored in the fœtal liver; the latter is essential for the maintenance of pregnancy in some animals, though not necessarily in man. The importance of vitamin K in the prevention and treatment of hæmorrhagic disease of the new-born is referred to on p. 149. The water-soluble vitamins traverse the placenta easily. All anæsthetics used to relieve pain at parturition pass through the placenta and if given in sufficiently high dosage to the mother they produce fœtal anæsthesia.

It can be concluded that in general substances of small molecular weight such as oxygen, carbon dioxide and perhaps glucose (mol. wt. 180) pass through the placental barrier in either direction by diffusion. Substances of higher molecular weight either do not pass at all (for example, sucrose, mol. wt. 342) or are transported in some special way in which numerous enzymes, notably phosphatase, are likely to be concerned.

THE FŒTUS

The human fœtus is very small up to the middle of pregnancy and consequently the nutritional demands from the mother's point of view are negligible. More than half the weight is acquired in the final six to eight weeks (Table 52, 23) and during this time, if the mother does not receive sufficient food, the fœtus draws on the maternal tissues. The birth weight of the infant is not, therefore, governed by the mother's diet unless there is severe undernutrition. Individual birth weights of the children of a multiple pregnancy are on the average lower than those of children of single birth. The mean birth weights of singletons, twins, triplets and quadruplets born in this country are respectively 3·38, 2·40, 1·82 and 1·40 kg. The rate of fœtal growth in multiple pregnancies is retarded in the last few weeks, perhaps partly because the smaller size of the placenta restricts fœtal growth.

Table 52, 23

Composition of the Human Fœtus in grams

	Body	Water	Protein	Fat	Ash
End of 5th lunar month . .	300	260	22	3·5	1·5
End of 7th lunar month . .	1000	800	100	25·0	30
End of 10th lunar month (birth) .	3200	2420	400	350·0	90
Average of daily deposition:					
1. Throughout fœtal life . .	11·4	8·6	1·4	1·25	0·32
2. In last three lunar months .	26·2	19·3	3·57	3·87	0·71
3. In last month . . .	35·7	23·6	6·4	5·0	2·0

A. St G. Huggett (1941), *Physiol. Rev.* **21**, 442.

By means of artificial insemination Walton and Hammond were able to cross a Shire mare with a Shetland stallion and a Shetland mare with a Shire stallion and to show that the birth weight of the progeny was determined by the size of the mare; even after birth the growth of the foal showed only a small paternal influence. There is evidence of a similar maternal factor determining the size of the offspring in human beings.

At least two other factors influencing birth weight are social class and prematurity. There is clear evidence that on the average the infant's birth weight is lower, the lower the social group to which the mother belongs; the reason for this is undoubtedly complex and the differences are not likely to be attributable to dietary causes alone. Any infant weighing less than 2·5 kg. (5 lb.) at birth is regarded as being premature and prematurity is associated with a higher death rate during the first month of extrauterine life (neonatal mortality). Furthermore a child whose birth weight is much below the average may for a long time be unable to reach the weight appropriate for his age, though this does not mean that he is weakly or mentally retarded.

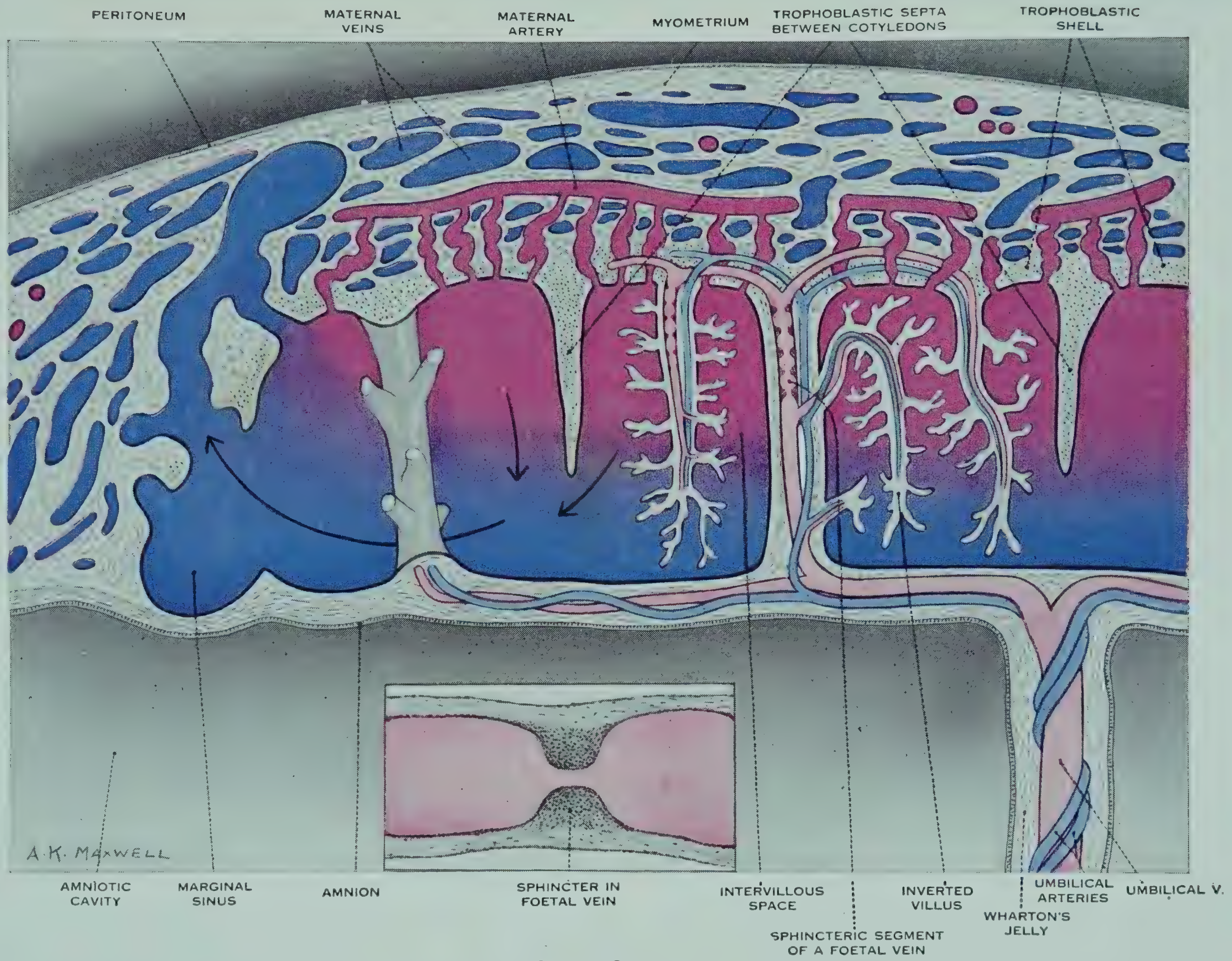

PERITONEUM
MATERNAL VEINS
MATERNAL ARTERY
MYOMETRIUM
TROPHOBLASTIC SEPTA BETWEEN COTYLEDONS
TROPHOBLASTIC SHELL
A.K. MAXWELL
AMNIOTIC CAVITY
MARGINAL SINUS
AMNION
SPHINCTER IN FOETAL VEIN
INTERVILLOUS SPACE
SPHINCTERIC SEGMENT OF A FOETAL VEIN
INVERTED VILLUS
WHARTON'S JELLY
UMBILICAL ARTERIES
UMBILICAL V.

FIG. 52, 24

Diagrammatic section of a portion of the placenta (*after* Spanner). (From W. J. Hamilton, J. D. Boyd & H. W. Mossman (1952), *Human Embryology*, Second Edition. Fig. 70, p. 73. Cambridge: Heffer.)

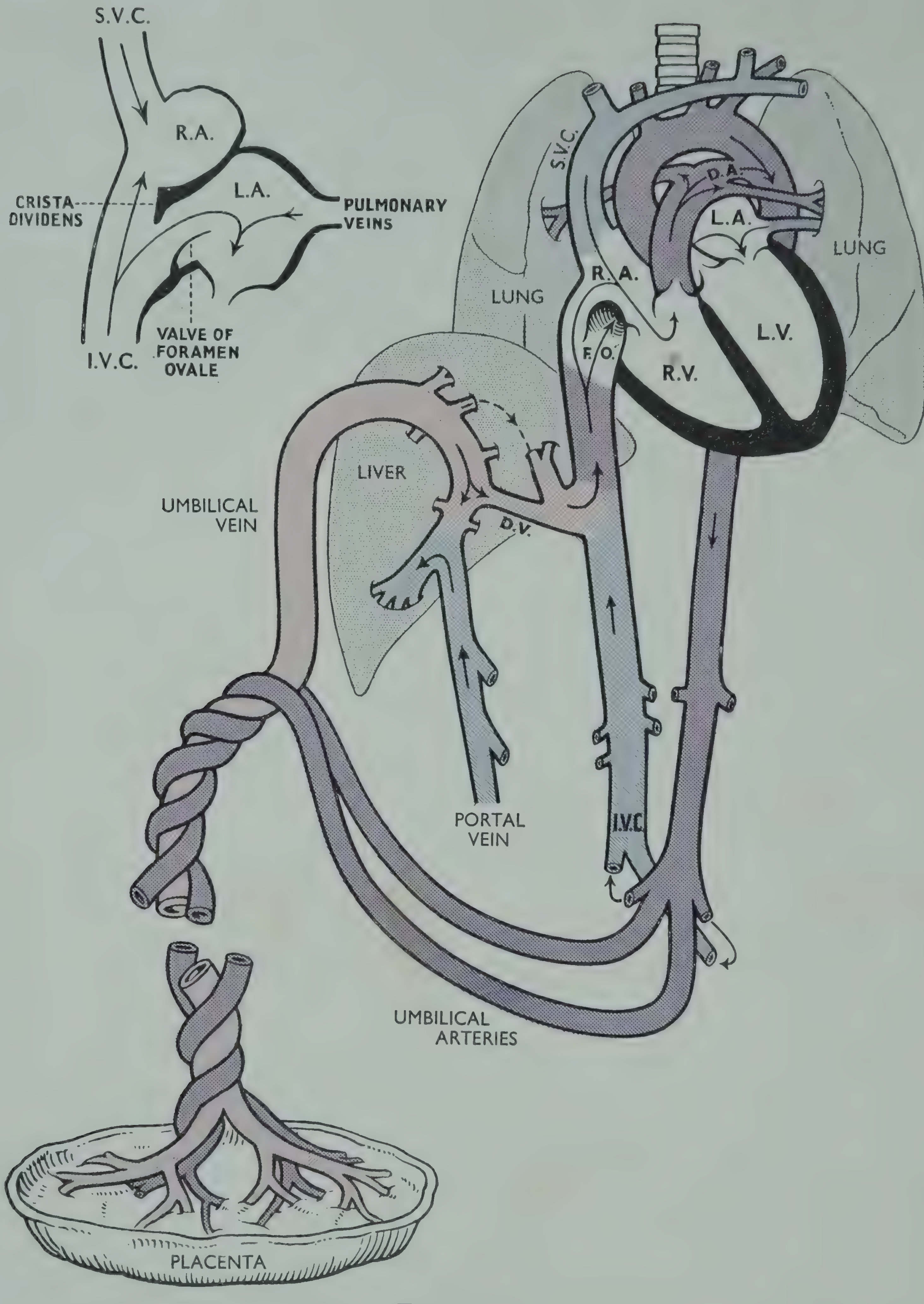

S.V.C.
R.A.
L.A.
CRISTA DIVIDENS
PULMONARY VEINS
I.V.C.
VALVE OF FORAMEN OVALE
S.V.C.
D.A.
L.A.
LUNG
LUNG
R.A.
L.V.
F.O.
R.V.
LIVER
UMBILICAL VEIN
D.V.
PORTAL VEIN
I.V.C.
UMBILICAL ARTERIES
PLACENTA

Fig. 52, 25

A plan of the fœtal circulation. The upper end of the inferior vena cava opens directly into the left atrium through the foramen ovale (see inset) as well as into the right atrium.

(*By courtesy of G. S. Dawes.*)

R.A. & R. V. = Right atrium and ventricle.
L.A. & L.V. = Left atrium and ventricle.
S.V.C. = Superior vena cava.
I.V.C. = Inferior vena cava.
D.A. = Ductus arteriosus.
D.V. = Ductus venosus.
F.O. = Foramen ovale.

Fœtal Circulation.—Until recently the course of the circulation in the embryo could be inferred only from dissections of human and animal fœtuses but it was shown experimentally by a group of workers at Oxford in 1939 that the classical descriptions of the fœtal circulation were substantially correct. They found the sheep the most convenient animal for their work because the long umbilical cord allows the fœtus to be delivered out of an incision in the uterus with the minimum of disturbance to the ewe or her offspring. If the mouth of the lamb is covered immediately after delivery by a rubber bag containing amniotic fluid, breathing of air is prevented. When the bag is removed from the muzzle of the lamb and the umbilical cord is tied, respiration begins and a number of circulatory adjustments takes place. The pathway taken by the circulating blood can be studied before and after respiration has begun by taking a number of X-ray photographs of the fœtus during the injection into the circulation of a radio-opaque material. The information afforded by this angiographic method has recently been supplemented in a quantitative fashion by Dawes and his colleagues at Oxford who have measured the oxygen saturation and the pressure of the blood at various points in the fœtal circulation (Fig. 52, 27).

The fœtal circulation (Fig. 52, 25 and 26 A) is quite different from that in the adult, chiefly because the heart has to pump the blood through the placenta as well as through the fœtal body. Of the blood passing down the descending aorta only a small part goes to the lower limbs, the major part passing by the two *umbilical arteries* to the placenta. After taking up oxygen there the blood is returned in the *umbilical vein* and flows either through the liver or by means of a by-pass, the *ductus venosus*, directly to the inferior vena cava which receives also the blood from the alimentary canal and from the lower extremities. At the cardiac end of the inferior vena cava the blood is divided by the edge of the interatrial septum, the crista dividens, into two streams, a large stream which goes directly through the foramen ovale into the left atrium and a somewhat smaller stream which reaches the right atrium. In the left atrium the blood coming from the inferior vena cava is mixed with a small pulmonary bloodstream and then passes to the left ventricle which pumps it to the arch of the aorta. This blood, the most highly oxygenated blood available in the systemic arteries of the fœtus, first supplies the heart by the coronary arteries and then the head and upper part of the body, the remainder reaching the descending aorta. The reduced blood returning from the head drains into the superior vena cava and right atrium where it is mixed with the reduced blood coming from the coronary sinus and with the right stream from the inferior vena cava which has a high oxygen content. The blood then passes to the right ventricle, the output of which is delivered partly into the lungs via the pulmonary artery but mainly to the aorta through the *ductus arteriosus*, a wide vessel about the same diameter as the aorta at the point where they meet. In this way the main stream of blood from the right ventricle by-passes the lungs and is distributed to the thorax and extremities as well as to the placenta through the umbilical arteries.

Fig. 52, 27 shows the percentage oxygen saturation of blood in eight of the principal vessels of the mature lamb before respiration is allowed to take place. From these figures the distribution of the blood flow can be calculated. The right atrium receives blood of 67 per cent. saturation from the inferior vena cava and of 25 per cent. from the superior vena cava and these streams are mixed and ejected into the pulmonary artery at a saturation of 52 per cent. This means that nearly two parts of inferior caval blood are mixed with one part of superior caval blood. On the other hand the left atrium receives inferior caval blood and pulmonary vein blood in the ratio of 4·5 to 1. The greater proportion of blood flow from the inferior vena cava into the left side of the heart explains the observation that the oxygen content of the brachiocephalic artery supplying the head is larger than that of the umbilical artery. Other calculations show that in the mature fœtal lamb the output of the left ventricle is 20 per cent. greater than that of the right. Inspection of Fig. 52, 27A shows that, as the two

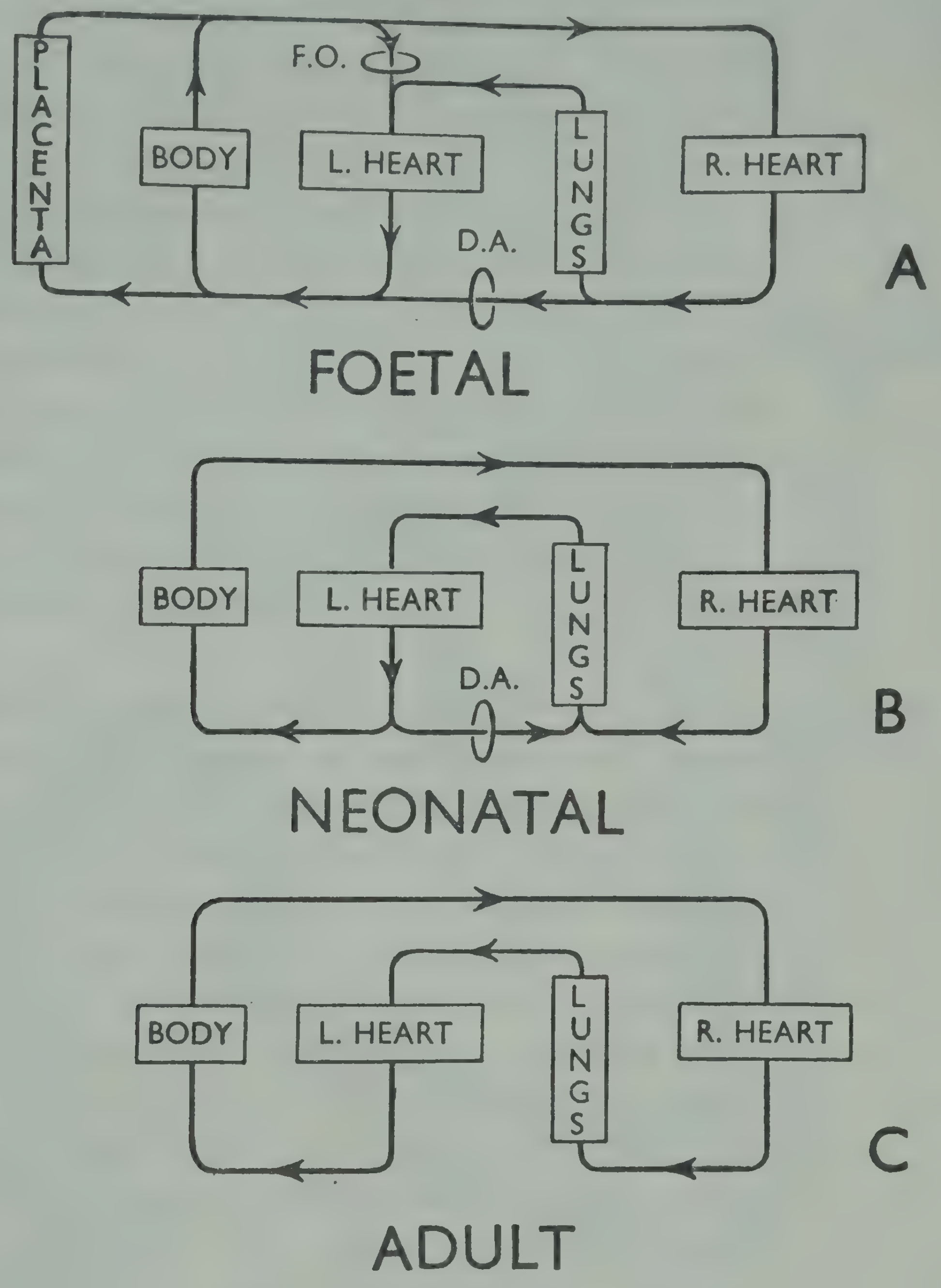


FIG. 52, 26

Much simplified diagrams to illustrate the changes in the circulation at birth as described in the text. In A the two ventricles are working in parallel to drive blood from the great veins to the arteries. B is the condition reached a few minutes after birth when the cord is tied and the foramen ovale F.O. closes. When the ductus arteriosus D.A. finally closes the adult circulation C is established with the two ventricles working in series. (G. V. Born, G. S. Dawes, J. C. Mott & J. G. Widdicombe (1954) *Cold Spr. Harb. Symp. quant. Biol.* **19,** 106.)

ventricles are not in series (as they are in the adult, Fig. 52, 27C), there is no need for their outputs to be equal. There is little difference in the thicknesses of the walls of the two ventricles in the fœtus and presumably little difference in the work done by them. The flow through the various tissues as a percentage of the combined output is, lungs, 10; other fœtal tissues, 35; placenta, 55; foramen ovale, ductus arteriosus and aortic isthmus, 40.

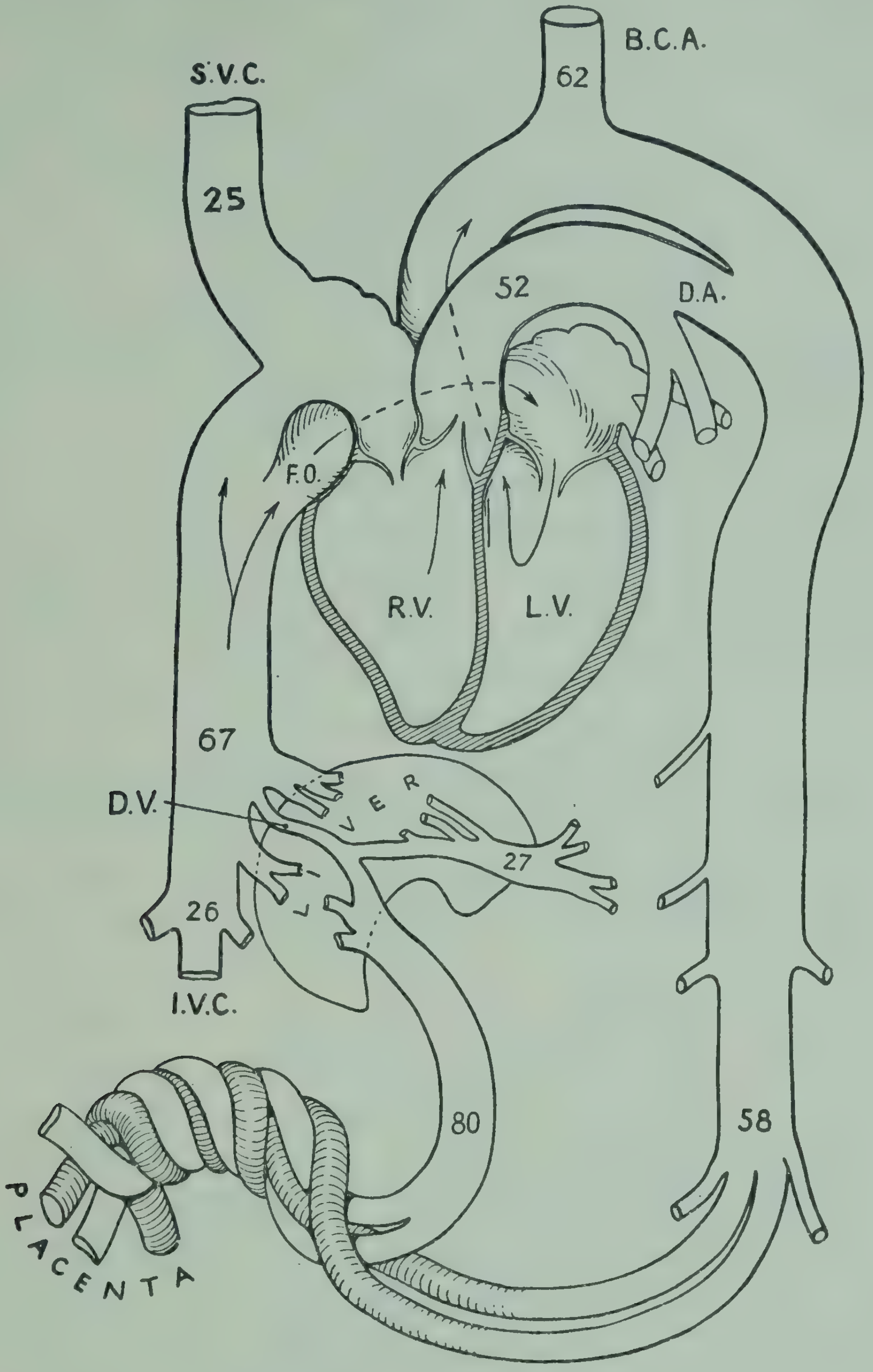


FIG. 52, 27

Diagram of the fœtal circulation in the lamb. The figures indicate the percentage oxygen saturation of blood withdrawn simultaneously from various vessels and averaged from determinations on six lambs. I.V.C., inferior vena cava; S.V.C., superior vena cava; D.V., ductus venosus; F.O., foramen ovale; D.A., ductus arteriosus; B.C.A., brachio-cephalic artery. (G. V. R. Born, G. S. Dawes, J. C. Mott & J. G. Widdicombe (1954), *Cold Spr. Harb. Symp. quant. Biol.* **19**, 103.)

When the mother is made to breathe air containing a low percentage of oxygen the fœtal oxygen saturation falls; fœtal heart rate and blood pressure increase and umbilical blood flow also increases. Since the increased umbilical flow and rise of blood pressure can be produced by injecting adrenaline or noradrenaline into the lamb, it is possible that the mechanism compensating for maternal asphyxia is a reflex release of sympathetic amines originating by stimulation of the chemoreceptors of the carotid and aortic bodies. When the fœtus is very severely asphyxiated and the arterial oxygen saturation falls below 15 to 20 per cent. the heart begins to slow; this is a useful indication of fœtal distress to the obstetrician.

At a normal birth the umbilical cord is torn as the animal is delivered, the umbilical arteries constrict. The umbilical vessels are extremely sensitive to trauma and contract firmly if, for example, an attempt is made to insert a needle into them to obtain blood. This is a local response of the smooth muscle and not a reflex adjustment since these vessels, which are outside the fœtal body, have no nerves.

At birth the cord is tied (or in animals is ruptured) and breathing begins. The sudden cessation of umbilical flow increases the peripheral resistance suddenly and causes a rise of fœtal blood pressure; at the same time the arterial oxygen content falls. When breathing begins the expansion of the lungs is associated with a great decrease in pulmonary vascular resistance; hence the rate of blood flow through the lungs increases five- to ten-fold within a minute or so. Consequently the pressure in the left atrium rises; the pressure in the inferior vena cava falls as the blood flow from the placenta is cut off and the valve of the foramen ovale (Fig. 52, 25) closes. Thus from within a minute of breathing all the venous blood enters the right atrium and passes through the lungs (Fig. 52, 26B). About a week after birth the valve of the foramen ovale becomes fused to the atrial wall. At the junction of the umbilical vein and the ductus venosus there is a small sphincter of smooth muscle which is responsible for closure of the ductus venosus after birth.

Before the animal breathes the pulmonary arterial pressure exceeds the pressure in the descending aorta. The decrease in pulmonary vascular resistance results in a fall in pulmonary pressure and in a reversal of the direction of blood flow through the ductus arteriosus (Fig. 52, 26 B); that is to say within a few minutes of birth blood flows from the aorta to the pulmonary artery and may contribute half or more of the total pulmonary blood flow. This stream of blood flowing through the left heart, the ductus arteriosus and the lungs increases the pulmonary blood flow and leads to a substantial increase of carotid oxygen saturation; temporary occlusion of the ductus at this stage causes a fall of systemic arterial oxygen saturation. This increased flow is especially important at this stage for it must be remembered that at this time the lungs are not properly expanded, nor have they cleared themselves of the amniotic fluid with which they were filled in utero. Closure of the ductus arteriosus begins within the first ten minutes but it is not complete for many hours. In the lamb the external diameter is reduced from 10 mm to 5 mm in several minutes but as the aortic pressure is rising at this time and the pulmonary pressure is falling, (Fig. 52, 28) a large amount of blood continues to flow through with the production of vibrations which can be heard as a murmur. This reverse flow through the ductus arteriosus from aorta to pulmonary artery continues for many hours after birth. At this time the left ventricular output is greater than the right. The circulation is now at the stage represented in Fig. 52, 26B. The flow of blood through the ductus ceases after two days or so and then the adult type of circulation is reached (Fig. 52, 26C) in which the ventricular outputs are equal. The closure of the ductus arteriosus and ductus venosus is finally made permanent by a proliferation of the lining endothelium which takes several months to complete and eventually the obliterated vessels are represented only by fibrous cords.

It is likely that the circulatory changes at birth in the human fœtus are very similar. Although the umbilical cord continues to pulsate for some time after the child is delivered it is customary to tie the cord as soon as the child is born. This may not, however, be the best practice because if drainage of blood from the placenta to the child is allowed it may gain 30 to 60 ml. of blood and so increase its stores of iron.

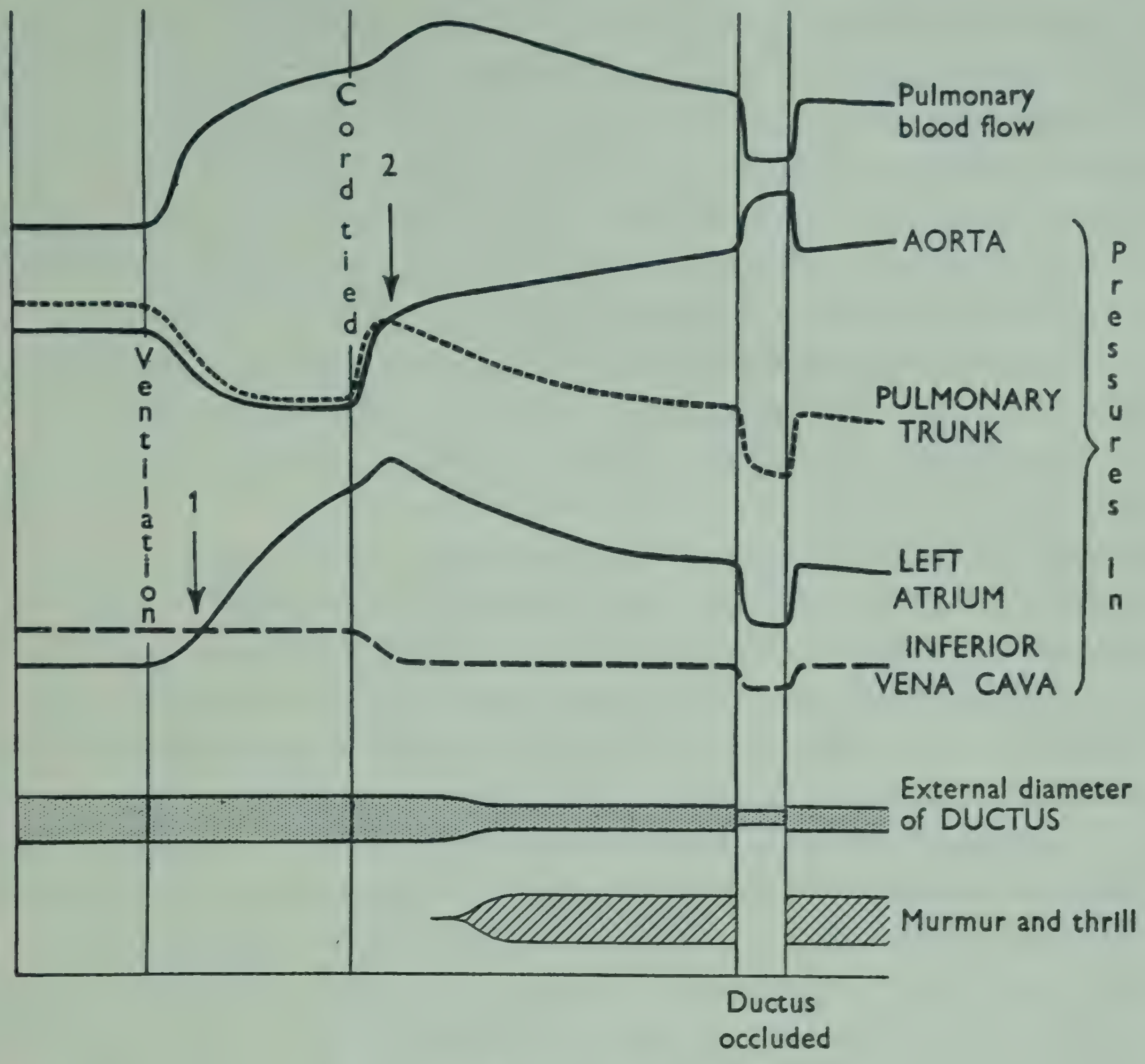


FIG. 52, 28

Diagram to illustrate the changes which occur in the circulation of a lamb when ventilation is begun before the umbilical cord is tied. At 1 the mean pressure in the left atrium rises above that in the inferior vena cava and thereafter the valve of the foramen ovale is closed. At 2 the pressure in the pulmonary trunk falls below that in the descending aorta and the direction of blood flow through the ductus arteriosus is reversed. Temporary occlusion of the ductus arteriosus now reduces pulmonary blood flow and abolishes the murmur and thrill which have appeared in the pulmonary trunk. (G. V. Born, G. S. Dawes, J. C. Mott & J. G. Widdicombe (1954), *Cold Spr. Harb. Symp. quant. Biol.* **19**, 104.)

PARTURITION

The hormonal pattern of pregnancy is so different in different mammalian species that it is impossible at present to give a single explanation of the onset of labour that can apply to all mammals. The problem is further complicated by the technical difficulties of obtaining data in the human subject and it is therefore possible to provide only a somewhat inadequate account of the processes involved.

Labour occurs in woman at a time equivalent to ten menstrual cycles, or 280 days, from the first day of the last menstruation. There is no evidence that the length of the menstrual cycle has an influence on the duration of pregnancy. It is well known that the period of gestation cannot be predicted exactly but if a margin of seven days is allowed on either side then a forecast of the date of delivery will be correct in 50 per cent. of cases. Since the non-pregnant uterus shows spontaneous activity (p. 959) it is not surprising to find that contractions occur early in pregnancy; they are small and usually separated by periods of an hour or so in which the uterus is quiescent. From the twentieth week the contractions are greater in amplitude and much more frequent, although they remain irregular until the last few weeks of pregnancy when rhythmic contractions begin. The interval between the contraction waves in late pregnancy is usually from five to ten minutes. The contractions which deliver the child at full term may be regarded as an exaggeration of this spontaneous uterine activity.

At the end of pregnancy in woman there is a reduction in the pregnanediol excretion (Fig. 52, 21) and presumably a fall in the production of progesterone by placenta and ovaries, for which no reason has yet been found. A similar fall in pregnanediol excretion is encountered in abortion occurring in the early stages of pregnancy and a threatened abortion has occasionally been prevented from becoming complete by the subcutaneous injection of progesterone. The success of this treatment has often been attributed to an inhibitory effect of progesterone on uterine activity but, as mentioned earlier (p. 979), it is difficult to demonstrate such an action. Just before parturition a large amount of conjugated œstrogen is found in the urine of the pregnant woman (Fig. 52, 20) and there is a rise in the blood œstrogens which might account for the increase in the reactivity of the uterus to oxytocin occurring at the end of pregnancy. It is difficult, however, to believe that this is of much significance in determining the onset of parturition, especially as attempts to bring about labour before full term in the human subject by giving large doses of œstrogens are usually unsuccessful.

At the thirty-second week the human uterus begins to react to oxytocin from the neurohypophysis (p. 1009). Injections of pituitary extracts are often given to stimulate uterine activity when it is desired to bring on labour a little before the expected time and they are also used to make the uterus contract firmly after delivery so as to reduce hæmorrhage. Although the reactivity of the uterus to oxytocin in several species, including man, increases up to full term this does not prove that oxytocin is the hormonal factor responsible for parturition. If it is, one would expect that hypophysectomy would delay parturition but this is not necessarily the case. In animals removal of the infundibular process does not necessarily lead to delay or difficulty in delivery but this might be due to the secretion of oxytocin by the stalk and median eminence which are left intact in this operation. Sometimes in cases of diabetes insipidus, in which the neurohypophysis is presumably damaged, labour has been prolonged. Although there is evidence that an oxytocic hormone is liberated from the neurohypophysis (p. 1009) it has not yet proved possible to obtain direct confirmation of the oxytocic theory of parturition.

It was shown by Sir James Simpson in 1856 that the nervous system is not responsible for the uterine contractions at parturition. In his experiments, destruction of the thoracic and lumbar cord in pigs did not interfere greatly with delivery although the contractions of the abdominal muscles were abolished. These muscles, which normally aid the uterus to expel the fœtus, are, of course, under the control of the spinal cord and higher centres. There is evidence that near full term the onset of labour may be hastened by emotional factors but it is not affected by section of the sympathetic and parasympathetic nerves.

Since the ovaries can be removed early in pregnancy in women without causing abortion and since the corpus luteum begins to degenerate in the third month of pregnancy, the hormones produced in the ovary can have little influence in determining the onset of parturition. The placenta certainly produces gonadotrophic hormones and probably elaborates œstrogens and progesterone, all of which are apparently necessary for the maintenance of pregnancy. In some animals, for example the mouse, the fœtuses can be removed without bringing the pregnancy to an end immediately and the placentæ are delivered some days later at full term. In other animals pregnancy is terminated by removal of the fœtus. Distension of the uterus as a factor in the termination of pregnancy is difficult to assess since in women the duration of pregnancy remains constant in spite of great variations in size of the fœtus and in spite of the presence of excessive amounts of amniotic fluid (hydramnios); even in extrauterine pregnancies in which the uterus is small and empty and the embryo is in the uterine tube or abdominal cavity the patient may experience pains at full term. Furthermore single pregnancies come to an end at a distension which is much less than that obtaining in a multiple pregnancy long before parturition. If distension of the uterus were the only exciting factor in parturition then it would be expected that it would occur when a certain total litter weight was reached. This is not so since the total litter weights in singletons, twins, triplets and quadruplets at birth are 3·38, 4·79, 5·45 and 5·58 kg. respectively. It is likely however that the uterus becomes less tolerant to distension as pregnancy advances since the mean period of gestation decreases in the four categories just given in the order 280·5, 261·6, 246·8 and 236·8 days.

The muscular activity of the human uterus can be studied by passing a catheter through the cervix, or through the abdominal wall and through the uterine wall, into the amniotic cavity and by connecting it to a manometer which records the variations in pressure. The slight contractions of the uterus which occur throughout pregnancy by increasing the pressure in the uterine veins assist the movement of blood through the intervillous placental spaces and so aid the oxygenation of the fœtal blood. Towards the end of pregnancy, but long before labour begins, the basal amniotic pressure is about 5 mm Hg with slow oscillations of 1 to 5 mm Hg above this value. At the onset of labour these rhythmic contractions increase and the basal intrauterine pressure rises to 8 to 12 mm Hg. In the *first stage* of normal labour the intrauterine pressure rises slowly with each contraction to a peak of 30 to 50 mm Hg above the basal value and then more slowly falls to the basal level. These contractions, which occur at intervals of two to five minutes, can be felt by the hand placed on the abdominal wall. When the pressure is above a certain value, which varies with different patients, but is in the region 25 to 35 mm Hg, the subject is conscious of a labour pain, that is a pain commencing in the lumbosacral region and radiating to the front and down the thighs. By pressing on the amniotic fluid the uterine contractions dilate first the upper part of the cervical canal and later the os uteri. The membranes finally rupture and a small quantity of amniotic fluid escapes. In the *second stage* of labour the intrauterine pressure at the height of a contraction may rise to a maximum of 110 mm Hg and the pressure between contractions is usually about twice that recorded in the first stage. The head of the child is slowly forced through the pelvis and is born, the remainder of the body following almost at once. During this stage voluntary contraction of the abdominal muscles can double the intrauterine pressure, so that the combined pressures caused by the contractions of the uterine and of the abdominal muscles may be as much as 260 mm Hg. It has been calculated that the force on the fœtal head produced by the uterine contractions alone is between eight and eighteen pounds, but in well-developed women the abdominal muscles can increase the expulsive force up to forty pounds or more.

Alvarez and Caldeyro-Barcia of Montevideo have recorded uterine activity by placing small balloons (0·02 ml.) into the myometrium at various positions. When the muscle contracts the pressure rises in the balloons and is recorded. They find that the contraction wave in labour starts at a pacemaker at the uterine end of one uterine tube, either right or left; the wave is propagated mainly downwards to the cervix at a speed of one or two cm per sec and reaches its peak in 50 sec or thereby. The contraction lasts longer in the parts near the pacemaker and diminishes in intensity as it passes down the uterus.

The pressure within the uterus during a contraction is sufficient to squeeze out the maternal blood from the uterine vessels into the vena cava and even occasionally into the aorta if the intrauterine pressure rises above the maternal arterial pressure. Thus at the height of each contraction the venous pressure is considerably raised, and this may be dangerous in patients with cardiac disease. There is a slight increase in arterial pressure during a labour contraction and a marked increase if the abdominal muscles contract at the same time.

After the birth of the child uterine contractions cease for a time, usually ten to fifteen minutes, and then in the *third stage* of labour they begin again and the placenta and membranes are expelled. With the detachment and delivery of the placenta a raw bleeding surface is left in the uterus which, however, contracts down so firmly that the venous sinuses are closed and usually little blood is lost. When the child is allowed to suck, there is an increase in uterine contractions which may be of value in preventing hæmorrhage. Hæmostasis is secured by coagulation of the blood in the uterine vessels which are later closed by fibrous tissue. During the *puerperium*, that is the period immediately following parturition, the uterus shrinks until, when involution is complete, it is only a little larger than in the virgin state.

THE MAMMARY GLAND

The mammary gland consists of numerous lobules made up of clusters of rounded alveoli embedded in fatty tissue. The secretion of the cells lining the alveoli passes into small ducts, and these unite with ducts from neighbouring alveoli to form the lactiferous tubules which radiate towards the mammary papilla or nipple and open on its summit. Beneath the areola, just before it reaches the nipple, each tubule has a dilatation, the ampulla, which acts as a milk reservoir. In the non-pregnant woman the alveoli are small but during pregnancy they enlarge and the cells lining them increase in number.

Both sexes possess mammary glands but they remain rudimentary in the male. In the girl enlargement of the breasts is one of the signs of puberty. In many animals there is an enlargement of the glands when corpora lutea are present in the ovaries; in women, although there may be a transient slight enlargement during the luteal phase of the menstrual cycle, full development is achieved only as a result of pregnancy. In a non-pregnant woman an increase in the size of the mammary gland can be produced by injection of œstrogen but a greater effect can be obtained by treatment with œstrogen plus progesterone. While œstrogen seems to be responsible chiefly for growth of the ducts, the combination of œstrogen and progesterone causes, in addition, alveolar growth. Because œstrogens are less effective in hypophysectomized animals it is possible that their action on the mammary gland is, in part at any rate, an indirect one exercised through the pituitary.

It has been claimed that a specific hormone, mammogen, from the pituitary is necessary for the full development of the mammary glands seen in pregnancy, but in the hypophysectomized spayed rat full mammary development can be obtained with two purified anterior pituitary hormones, prolactin and somatotrophin, together with œstogen and progesterone. In this animal and probably other species there is no need to postulate on special mammogenic hormone. There is no doubt that in cows somatotrophin stimulates lactation. The placenta may also influence the mammæ through its output of œstrogens and progesterone. Although during pregnancy it is often possible by pressure to expel a few drops of milk from the breasts, a free flow of milk does not occur until the second or third day after the birth of the child. It seems that during pregnancy the mammary glands are kept in an unresponsive state by the direct action of the large amount of œstrogens and progesterone in the circulation. The decline in the amount of these hormones in the blood after parturition allows the lactogenic hormone, *prolactin* (LTH), from the anterior pituitary to act directly on the alveolar epithelium. The amount of LTH in the blood and pituitary is not altered by the rise in œstrogens. Prolactin may not be the only lactogenic hormone secreted by the adenohypophysis and it is likely that adrenocorticotrophin supports lactation in an indirect way by keeping up the supply of milk precursors. The lactogenic hormones cannot bring about milk secretion unless the mammary gland has been prepared by the ovarian hormones. During the puerperium the administration of large amounts of either natural or synthetic œstrogens can stop the flow of milk by a direct action on the breast tissue.

Although the building up of secreting breast tissue and the initiation of lactation is mainly a hormonal affair there is no doubt that regular removal of milk from the breast is essential to maintain the flow. A woman with a strong, healthy, hungry child who sucks vigorously, secretes more milk than one with a small weak child. In dairy cattle more frequent milking produces more milk. Suckling may act by increasing the production of a galactopoietic hormone from the adenohypophysis; the hormone responsible may be somatotrophin but is not prolactin or ACTH. The release of the hormone may be brought about by reflex stimulation of the hypothalamus which stimulates the adenohypophysis by a humoral mechanism (see p. 1001). The maintenance of lactation requires also an adequate supply of adrenal cortical hormones and also thyroid hormone.

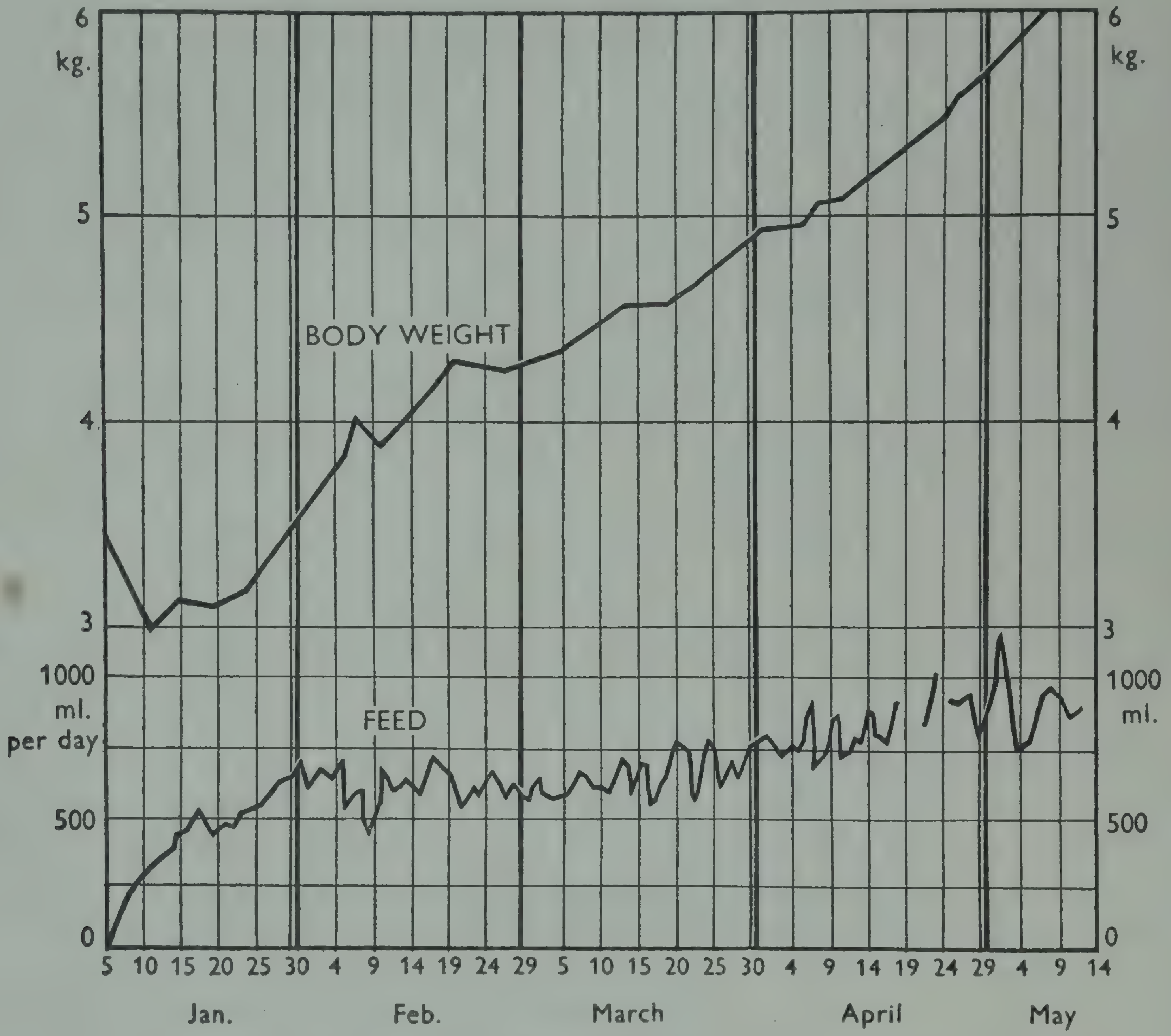


FIG. 52, 29

Weight and test feeding chart of a healthy breast-fed baby. The amount of milk obtained at each feed is measured by weighing the baby before and after suckling (*After* P. Evans (1950), *Brit. med. J.* i, 1128).

Although it is well known that lactation can be influenced by the emotional state of the mother it has not been possible to demonstrate experimentally the existence of a nervous control. Sympathectomy does not affect milk production and no one has demonstrated secretory nerve fibres in the mammary gland. Lactation cannot be fully successful unless the nursing mother is having a generous diet and an indication of the daily milk production in such circumstances is given in Fig. 52, 29.

It has been known for a long time that when a cow is milked there is a rise of pressure in the udder which aids the expulsion of the milk. This phenomenon, 'let-down' or the 'draught', was presumed by Hammond to be a reflex resulting from the stimulation of the teat. In the rabbit suckling causes, after a latent period of ½ to 1½ min, a rapid rise of intramammary pressure which transfers the milk rapidly to the offspring sucking at the teats. The functional importance of this reflex is shown by the fact that if it is abolished by anæsthesia or by emotional disturbances the amount of milk obtained by the young is greatly reduced. The 'let-down' phenomenon has nothing to do with production of milk but merely with the ejection of milk which is already formed in the gland. The mechanism can, however, be only partly nervous since cutting the nerve supply to the gland does not impair milk ejection. A preliminary clue to the solution of the problem was provided by the finding that injection of small doses of oxytocic extracts of the neurohypophysis produces ejection of milk in anæsthetized does in a manner closely resembling the natural phenomenon (Fig. 52, 30). That this is in fact the basis of the effect has been demonstrated quite recently. Electrical stimulation of the infundibular stem, paraventricular nucleus, supraoptic nucleus, median eminence or the neural lobe itself in anæsthetized does causes ejection of milk after a latent period of 20 to 45 sec; such a long latency is characteristic not of a reflex but of a humoral effect and is assumed to be the time taken for

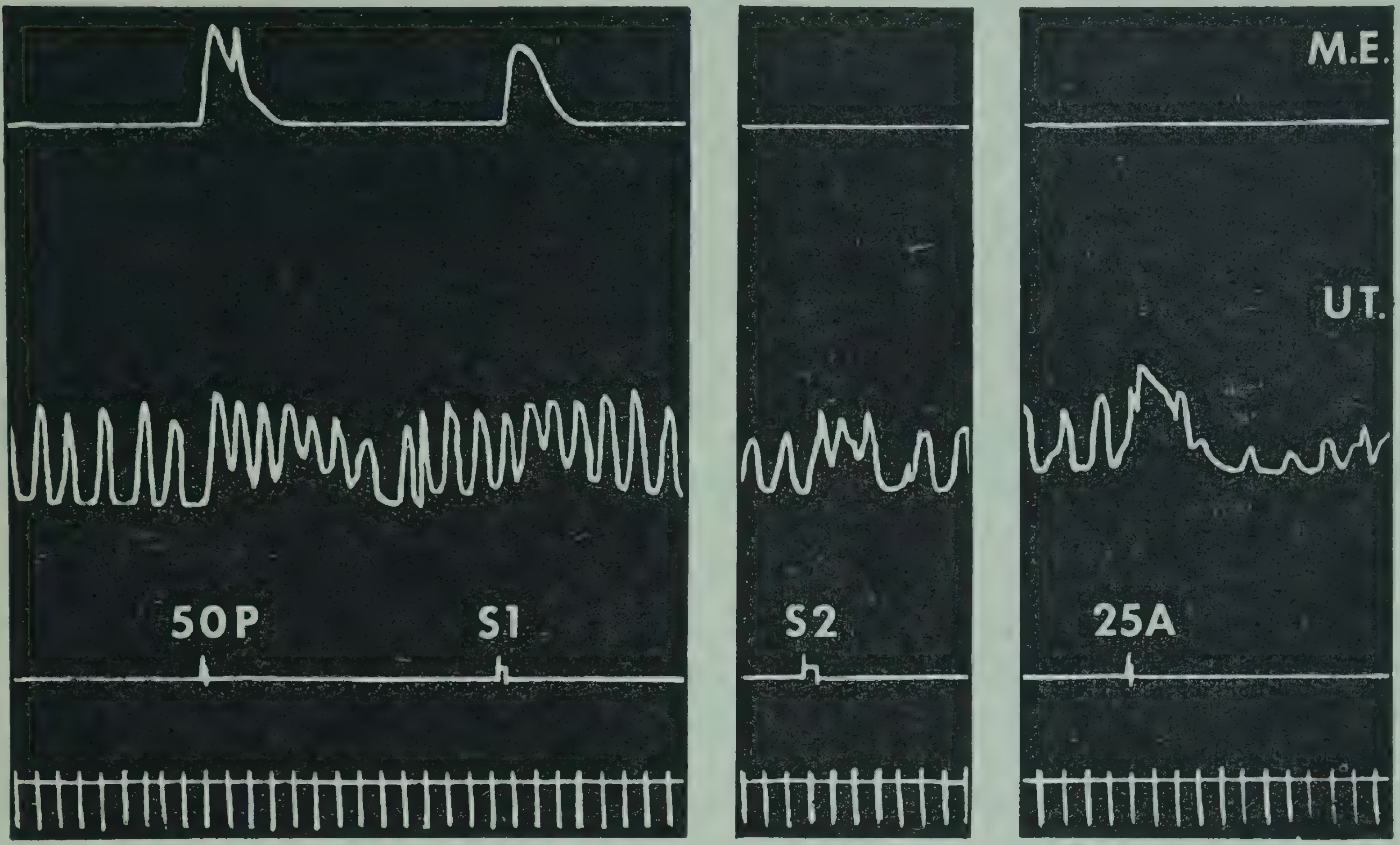


Fig. 52, 30

Effect of hypothalamic stimulation on milk ejection and uterine motility in an anæsthetized rabbit six days post partum.

M.E.: kymographic record of milk-ejection responses from cannulated teats.
U.T.: kymographic tracing of uterine activity recorded by means of a small intra-uterine balloon.
50P: 50 milliunits Pitocin (oxytocin) injected.
S1: 10-sec electrical stimulation of supraopticohypophysial (SOH) tract in infundibular stem.
S2: 15-sec electrical stimulation in right lateral hypothalamic area (pupil dilatation and hyperpnœa were observed).
25A: 25 μg. adrenaline injected.
Time signal registers 30-sec intervals.

Note 'spontaneous' uterine contractions, and similarity of the effects on the uterus and mammary gland of SOH stimulation and injection of Pitocin. Stimulation of the hypothalamic sympathetic area gave no milk ejection but evoked a uterine response similar to that following adrenaline. The response to adrenaline is followed by a partial inhibition of the rhythmic uterine activity.

(B. A. Cross (1955), *Brit. med. Bull.* **11**, 151.)

release of the hormone and for its transport to the mammary gland. After removal of the neurohypophysis electrical stimulation of these regions is ineffective and the young are unable to obtain sufficient milk to survive. If a lactating rabbit is forcibly restrained, the milk yield obtained by the young is reduced in some experiments to one-fifth. This reduction in most cases is due to a central block of the release of oxytocin since an intravenous injection of oxytocin restores milk ejection. In a few cases the inhibition may be due to the release of adrenaline in response to the emotional stimulus. If vasoconstriction is produced in a lactating mammary gland by stimulation of the nerves to it, or by injection of adrenaline, oxytocin can not cause milk ejection. The vasoconstriction blocks the access of oxytocin and does not put the milk ejection mechanism itself out of action; even after injection of adrenaline oxytocin applied directly to the mammary gland causes a local contraction of the myoepithelial cells surrounding the alveoli—these cells also still respond to mechanical stimulation. Whether the emotional reduction of milk ejection in women is due to one or other of these possible mechanisms has still to be discovered.

The fluid produced in the breasts during the first two or three days of the puerperium is called *colostrum*. It is small in amount and has a high protein and a low fat and sugar content. As the baby receives very few calories during this time he usually loses weight (Fig. 52, 29). Unless he is given water he may become dehydrated. The intake of breast milk and the growth curve of a healthy baby are shown in Fig. 52, 29. When lactation begins on the fourth to the fifth day the central cells of the mammary alveoli are discharged into the fluid, where they appear as the so-called colostrum corpuscles.

Mammary Metabolism.—The R.Q. of actively secreting mammary tissue is usually over 1·0 due to the synthesis of fat from carbohydrate. The mammary gland is an efficient milk producer since, provided the maternal caloric requirements have already been fully met, about 60 per cent. of the energy value of additional food is recovered in the milk. In order to avoid negative nitrogen balance during lactation about 2 g. of protein per kg. of body weight per day must be supplied to the mother. Calcium is less well absorbed during lactation than in pregnancy and a negative calcium balance is common.

Venous blood from the udder of a lactating cow contains about 10 mg. less glucose per 100 ml. than the arterial blood. By applying the Fick principle (p. 449) the total blood flow through the gland can be calculated from the arteriovenous glucose difference and the output of sugar in the milk if it is assumed that the lactose of the milk is derived entirely from the glucose of the blood. Calculations based on this assumption give what appear to be very high values for the blood flow and it is, therefore, probable that, while glucose is the source of a large part of the lactose, some must be derived from lactic acid or amino-acids. The volume of blood which passes through the mammary gland is 400 to 500 times the volume of milk secreted. The fat in milk contains short chain fatty acids not found in the blood and it has been shown that these are produced locally in the mammary tissue from acetate and from glucose. Both globulin and amino-acids are taken up by the gland but the former seems to be the main source for the synthesis of caseinogen.

REFERENCES

ALLEN, E. (Ed.) (1939). *Sex and Internal Secretions.* London: Baillière, Tindall & Cox.

ALVAREZ, H. & CALDEYRO-BARCIA, R. (1954). The normal and abnormal contractile waves of the uterus during labour. *Gynæcologica,* **138,** 190–212.

BARCLAY, A. E., FRANKLIN, K. J. & PRITCHARD, M. M. L. (1944). *The Fœtal Circulation and Cardiovascular System and the Changes that they undergo at Birth.* Oxford: Blackwell.

BARCROFT, J. (1946). *Researches on Pre-natal Life.* Oxford: Blackwell.

BEACH, F. A. (1947). A review of physiological and psychological studies of sexual behaviour in mammals. *Physiol. Rev.* **27,** 240–307.

BELL, G. H. (1952). Abnormal uterine action in labour. *J. Obst. Gynæc. Brit. Emp.* **59,** 617–623.

BURROWS, H. (1949). *Biological Action of Sex Hormones.* 2nd ed. Cambridge University Press.

CHANG, M. C. & PINCUS, G. (1951). Physiology of fertilization in mammals. *Physiol. Rev.* **31,** 1–26.

CORNER, G. W. (1943). *The Hormones in Human Reproduction.* Oxford University Press.

CORNER, G. W. (1951). Our knowledge of the menstrual cycle, 1910–1950. *Lancet,* i, 919–923.

CORNER, G. W. (1952). The events of the primate ovarian cycle. *Brit. med. J.* ii, 403–409.

COWIE, A. T. (1950). *Recent Studies on the Endocrine Control of Mammary Development and Milk Secretion,* pp. 81–95. Colloque International du C.N.R.S., Strasbourg.

DORFMAN, R. I. & SHIPLEY, R. A. (1956). *Androgens: Biochemistry, Physiology and Clinical Significance.* London: Chapman & Hall.

FOLLEY, S. J. (1949). Biochemical aspects of mammary gland functions. *Biol. Rev.* **24,** 316–354.

FOLLEY, S. J. (1950). *Biochemical Aspects of the Hormonal Control of Lactation,* pp. 81–95. Colloque International du C.N.R.S., Strasbourg.

HUGGETT, A. ST G. (1941). The nutrition of the fœtus. *Physiol. Rev.* **21,** 438–462.

HUGGINS, C. (1945). The physiology of the prostate gland. *Physiol. Rev.* **25,** 281–295.

LUNDQUIST, F. (1949). Aspects of the Biochemistry of Human Semen. *Acta Physiol. Scand.* **19,** Supplement 66.

MANN, T. (1954). *The Biochemistry of Semen.* London: Methuen.

MANN, T. & LUTWAK-MANN, C. (1951). Secretory function of male accessory organs of reproduction in mammals. *Physiol. Rev.* **31,** 27–55.

PARKES, A. S. (1945). The adrenal-gonad relationship. *Physiol. Rev.* **25,** 203–254.

PARKES, A. S. (Ed.) (1955). Hormones in reproduction. *Brit. med. Bull.* **11,** 83–170.

POPJAK, G. (1951–52). The metabolism of fat in the mammary gland and fœtal tissues with reference to the application of isotopic tracers. *Nutr. Abst. Rev.* **21,** 535–553.

Reviews of the progress of dairy science: reproduction and lactation (1950). *J. dairy Res.* **17,** 338–342.

REYNOLDS, S. R. M. (1949). *Physiology of the Uterus.* 2nd ed. London: Hamish Hamilton.

REYNOLDS, S. R. M. (1951). Determinants of uterine growth and activity. *Physiol. Rev.* **31,** 244–273.

ROBSON, J. M. (1947). *Recent Advances in Sex and Reproductive Physiology.* London: Churchill.

Symposium on lactation (1947). *Brit. med. Bull.* **5,** 123–220.

Symposium on nutrition and fertility (1949). *Brit. J. Nutr.* **3,** 79–107.

The biochemistry of fertilization and the gametes (1951). *Biochem. Soc. Symp.* No. 7.

The mammalian fetus: physiological aspects of development (1954). *Cold Spr. Harb. Symp. quant. Biol.* **19,** 1–225.

CHAPTER 53

THE PITUITARY BODY

THE pituitary body or *hypophysis cerebri* is an ovoid structure, some 10 × 10 × 5 mm in size, which occupies the hypophysial fossa in the sella turcica of the sphenoid bone covered over by the diaphragma sellæ of the dura mater. It is attached by a short stalk, the infundibulum, to the base of the brain between the optic chiasma and the mamillary bodies. The weight of the human hypophysis is usually about 0·6 g. but its size varies with age, sex, weight and height; in pregnancy the gland becomes much enlarged. Since the hypophysis occupies the whole of the sella it is clear that a pituitary tumour can grow only by compressing structures within the sella and later only by pressing on the base of the brain itself. As the optic chiasma lies immediately over the diaphragma sellæ an enlarging pituitary body, by exerting pressure on the optic nerves, chiasma, or tracts, can produce defects in the field of vision.

Table 53, 1

Subdivisions of the Pituitary Body (Terminology recommended by the International Commission on Anatomical Nomenclature).

Adenohypophysis	Lobus glandularis	Pars distalis	Anterior lobe.
		Pars tuberalis	
		Pars intermedia	Posterior lobe.
Neurohypophysis	Lobus nervosus (neural lobe)	Processus infundibuli.	
	Infundibulum (neural stalk)	Pediculus infundibularis (infundibular stem).	
		Bulbus infundibularis.	
		Median eminence of the tuber cinereum.	

Table 53, 1, and Figs. 53, 2 and 4, describe a classification of the various parts of the hypophysis based upon its embryology and this subdivision is useful in terms both of morphology and of function.

The tissues of the adenohypophysis are derived from the somatic ectoderm of the buccopharyngeal pouch, Rathke's pouch, which grows upward towards the third ventricle of the developing brain. The neurohypophysis, on the other hand, has its origin in a prolongation from the floor of the diencephalon in the region of the tuber cinereum which grows downwards towards the base of the skull so that it eventually lies immediately behind the buccopharyngeal pouch. The line of demarcation between the two parts is represented in the human adult by a few isolated cystic spaces.

Adenohypophysis.—Although minor cytological differences can be recognized between the different parts of the adenohypophysis the cells of this part of the pituitary are of three main types, namely *chromophobe*, *acidophil* and *basophil* (Figs. 53, 5 and 53, 6). The cytoplasm of the chromophobe cells is devoid of granules whilst the cytoplasmic granules of the acidophil cells have a strong affinity for acid dyes and those of the basophil group for basic dyes. The presence of chromophilic granules is an index of functional activity and it seems probable that the granules represent either hormones or their inactive precursors. Current views of pituitary function make it necessary to suppose that at least six hormones are produced by two morphologically distinct types of cell. It is possible, however, that only one or two proteins are actually liberated from the pituitary into the blood and that the multiplicity of their effects is the result of their ability to stimulate different target organs, for example thyroid and gonads, to differing extents. In either case it is not difficult to understand how disorders of the pituitary body usually result in a complex disturbance of metabolism. The chromophobe cells, accounting for about 50 per cent. of the cells in a normal gland, are believed to be cells that have discharged their secretion; as granules of secretion re-accumulate the cells again become chromophilic. It has been shown that in the rat there are two distinct types of chromophobe cell one of which develops basophil granules, the other acidophil.

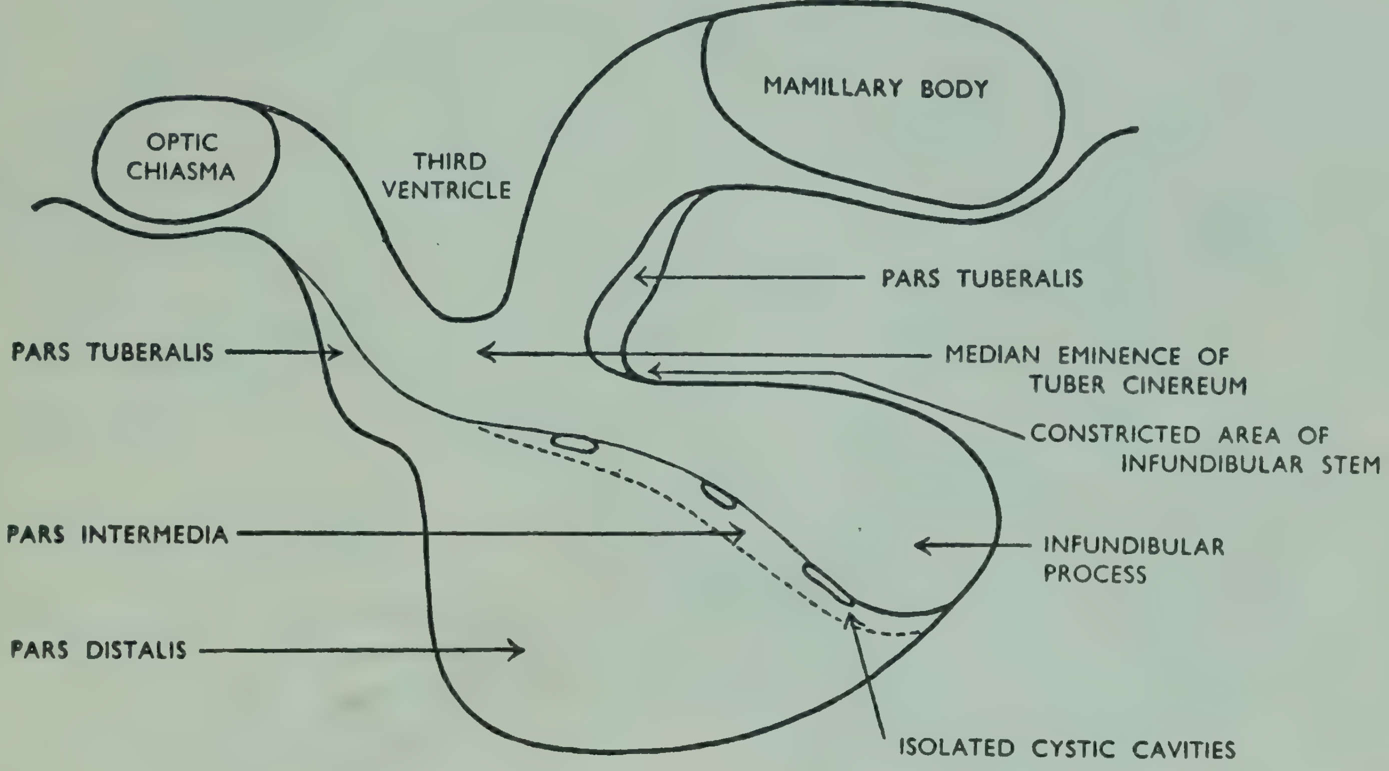


FIG. 53, 2

Diagram of sagittal section through the hypophysis and adjacent hypothalamus in adult man to illustrate the nomenclature. Comparison of this diagram with Fig. 53, 4 of the hypophysis of the dog shows differences in detail, for example the ventricular cavity does not extend in man into the infundibular process.

NEUROHYPOPHYSIS.—The cells of the neurohypophysis are of three kinds, ependymal cells, neuroglial cells and the so-called *pituicytes*, which in man form the greater part of the bulk of this part of the hypophysis. The last are large, granular, pyramidal, or spindle-shaped cells that have been credited with the elaboration of the hormones. Certain experiments, however, suggest that the active principles may actually be formed in cells of the supra-optic and para-ventricular nuclei and that they then pass down the nerve tracts to enter the blood in the infundibular process. The neurohypophysis contains also hyaline bodies (Herring bodies), the nature and function of which are unknown. It is doubtful if they contain hormones as was at one time suggested and it may be that they are artifacts.

NERVE SUPPLY.—The pituitary is richly endowed with nerves

(Fig. 53, 3). Postganglionic fibres from the superior cervical ganglia reach the adenohypophysis in the nerve plexus surrounding the internal carotid artery, which probably also conveys parasympathetic fibres from the petrosal nerves; there is no good evidence that either sympathetic or parasympathetic fibres have any regulating effect on the hormonal activity of the adenohypophysis. The neurohypophysis is connected with hypothalamic nuclei by a large number of fibres, estimated at 100,000 in man, collectively known as the hypothalamo-hypophysial tract, in which the main components arise in the supraoptic and paraventricular hypothalamic nuclei. The hypothalamo-hypophysial fibres end mainly in the neurohypophysis, only a few passing forwards to the pars tuberalis and pars intermedia as shown in Fig. 53, 3. The pars distalis does not receive any innervation from this source. After section of the hypothalamic fibres the neurohypophysis shrinks and ceases to function. The tuber cinereum and infundibulum are also connected with the hypothalamus by the tuberohypophysial tract, the origin and function of which are unknown.

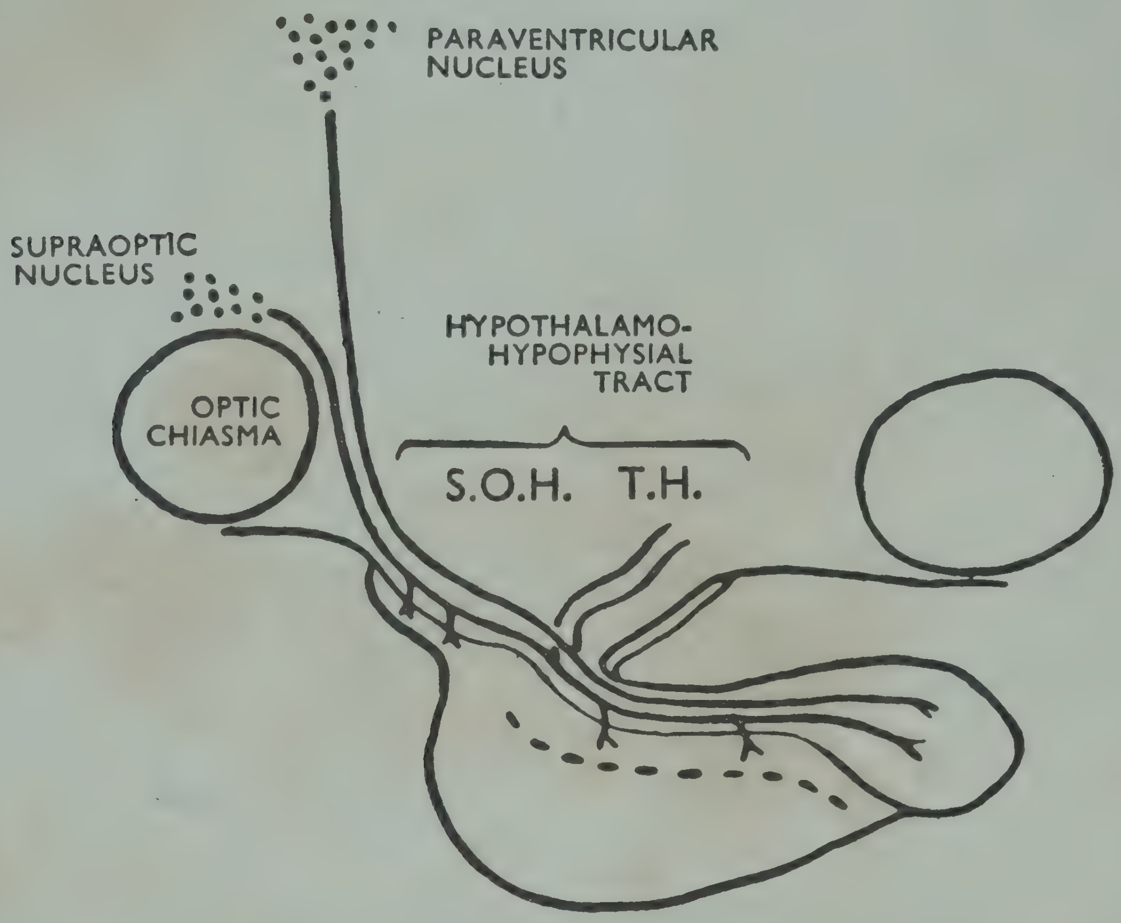


FIG. 53, 3

Diagram of a sagittal section through the pituitary gland. The hypothalamo-hypophysial tract, comprising the supraopticohypophysial (S.O.H.) and tuberohypophysial (T.H.) tracts, innervates the neurohypophysis. A few nerve fibres may be seen entering the pars tuberalis and pars intermedia, but it is doubtful if the pars distalis receives any innervation (G. W. Harris (1955), *Neural Control of the Pituitary Gland*. London: Arnold.

BLOOD SUPPLY.—The pars distalis and the infundibular process both receive blood directly from the internal carotid arteries by the anterior and posterior hypophysial branches. The circulation through the median eminence resembles that in the liver in that blood is conveyed by small twigs from the internal carotid arteries to a vascular plexus between the median eminence and the pars tuberalis. From this plexus, which is richly innervated, blood is drained into large portal trunks passing to the pars

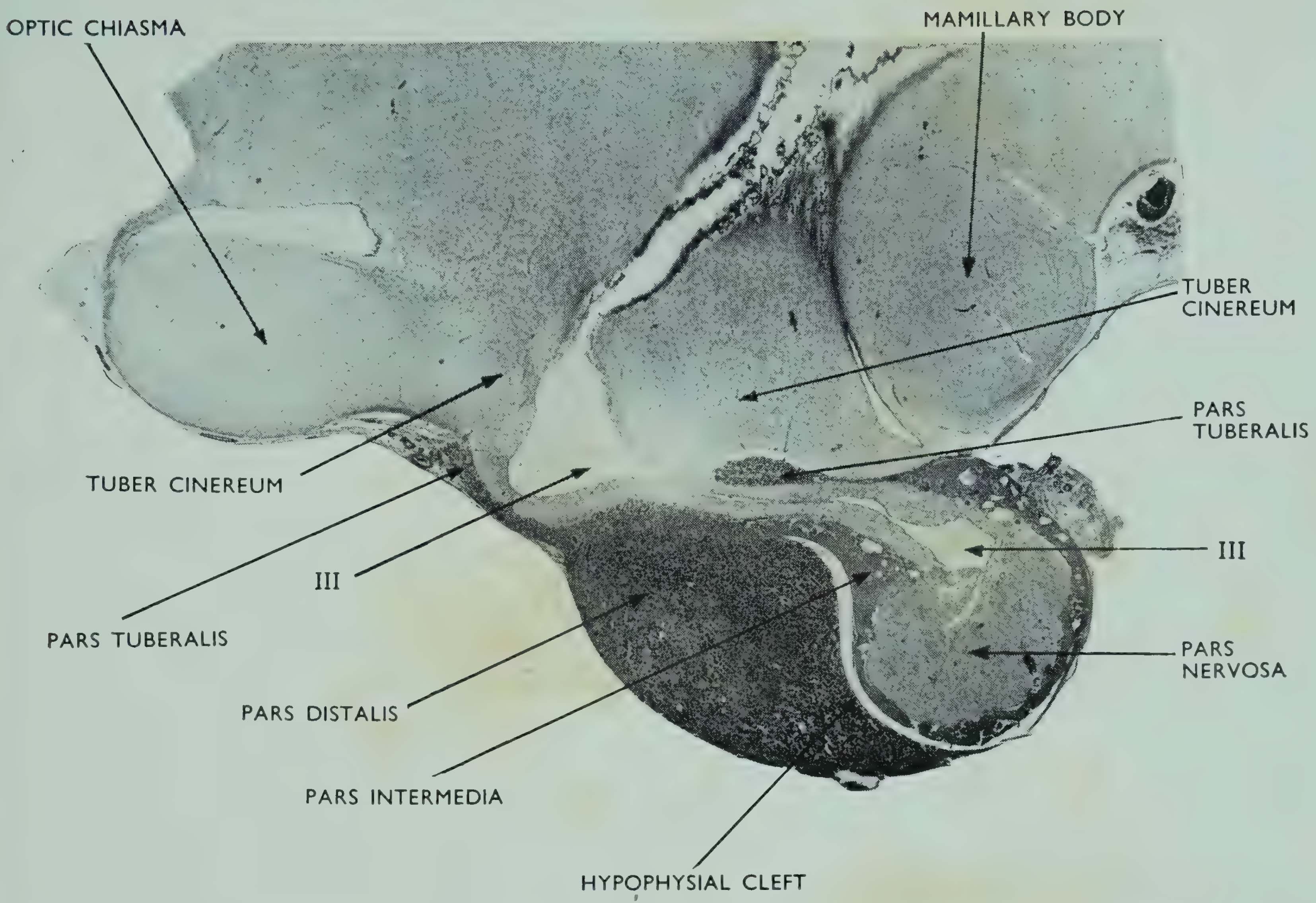

OPTIC CHIASMA
MAMILLARY BODY
TUBER CINEREUM
TUBER CINEREUM
PARS TUBERALIS
III
III
PARS TUBERALIS
PARS NERVOSA
PARS DISTALIS
PARS INTERMEDIA
HYPOPHYSIAL CLEFT

FIG. 53, 4

Mid-sagittal section through the hypothalmus and pituitary gland of the dog. III = Infundibular recess of third ventricle. Magnification × 10

(*By courtesy of K. C. Richardson.*)

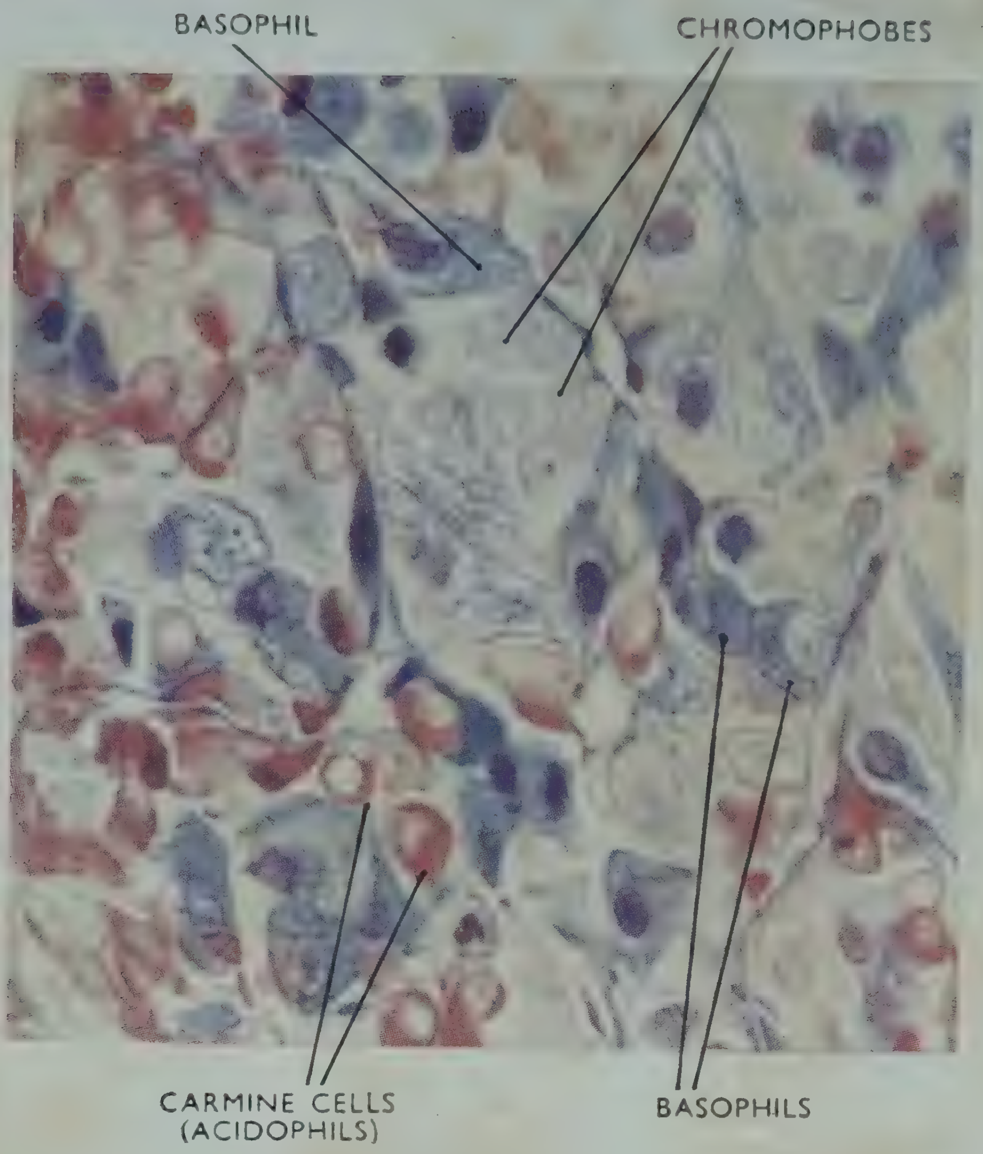

BASOPHIL
CHROMOPHOBES
CARMINE CELLS (ACIDOPHILS)
BASOPHILS

Fig. 53, 5

Rabbit adenohypophysis (*Heidenhain's Azan Stain*) 6½ hours after coitus. Carmine cells (acidophils)—cytoplasm red; regenerating basophils—cytoplasm blue; chromophobes—cytoplasm grey. (× 500.)

(*By courtesy of A. G. E. Pearse.*)

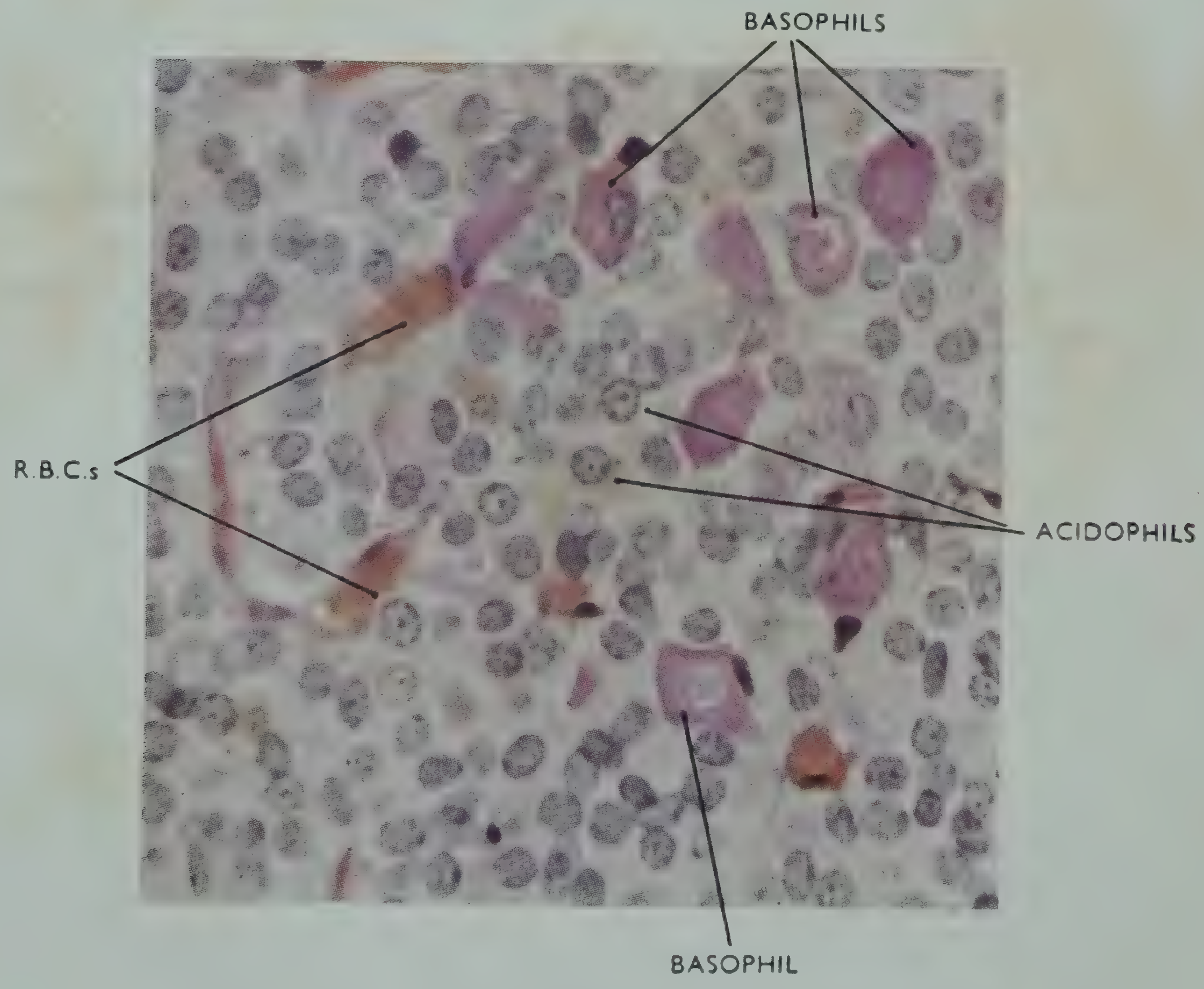

BASOPHILS
R.B.C.s
ACIDOPHILS
BASOPHIL

Fig. 53, 6

Rat adenohypophysis (*Tripas Stain—Pearse*). Mucoid (basophil) cells—cytoplasm red; acidophil cells—cytoplasm yellow; R.B.C.'s—yellow. (× 520.)

(*By courtesy of A. G. E. Pearse.*)

distalis, where they again break up into sinusoids (Fig. 53, 7). By this means substances produced in the hypothalamus are able to exert their actions on the pars distalis. The venous drainage from all parts of the gland is into the adjacent venous sinuses lying around the gland or in the underlying sphenoid bone. The hormones of the adenohypophysis presumably enter the circulation by this route. The hormones of the neurohypophysis are liberated under the control of the hypothalamo-hypophysial nerve fibres into the capillaries and veins of the neurohypophysis which is almost as vascular as the pars distalis.

The notion that the functional activity of the adenohypophysis is controlled through the hypophysial portal vessels has not received direct confirmation but many experiments make it likely that hypothalamic nerve fibres terminating around the multitude of capillary loops in the median eminence liberate some chemical substance into the blood by which it is conveyed to the pars distalis to regulate its secretory activity. The best evidence for this view is that grafts of the anterior pituitary into hypophysectomized rats, although they become vascularized and remain alive if implanted in other parts of the brain, do not become functional unless they are placed under the median eminence where they acquire vascular connexions with the hypophysial portal vessels. The successful function of such grafts can be judged by their ability to support normal œstrous cycles, pregnancy and lactation.

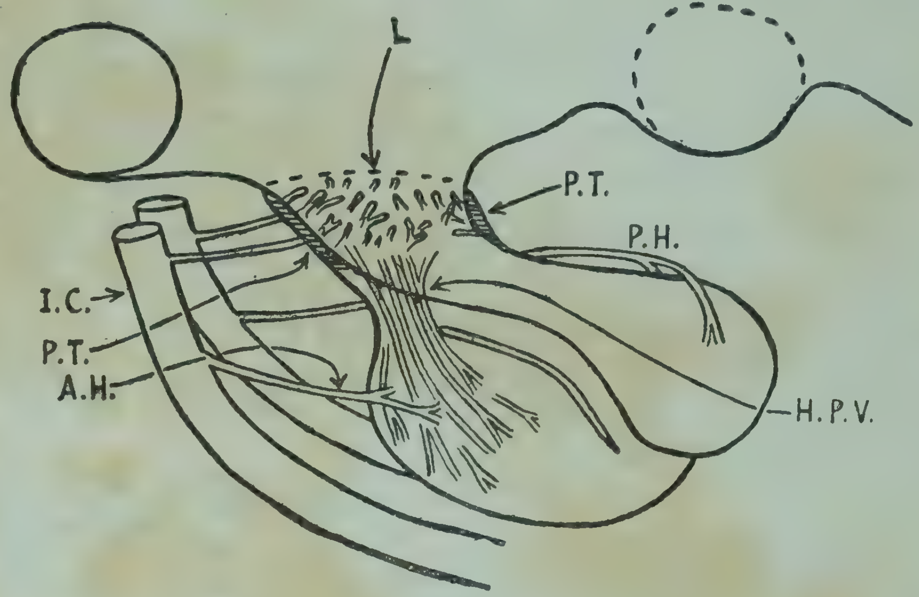


Fig. 53, 7

Diagram of a sagittal section through the pituitary and hypothalamus to illustrate the systemic arterial supply of the pituitary gland, and the hypophysial portal system. A.H. and P.H., the anterior and posterior hypophysial arteries; I.C., branches from the internal carotid arteries, supplying the pars distalis and infundibular process, respectively. Small twigs from the internal carotid arteries run to the pars tuberalis (P.T.) and supply a vascular plexus situated between this structure and the median eminence. From this plexus capillary sinusoids arise and penetrate the median eminence in the form of loops or tufts (L.). These capillaries unite into the large portal trunks (H.P.V.) and pass down into the pars distalis (G. W. Harris, (1948), *Brit. med. J.* i, 339).

The nature of the humoral transmitter (or transmitters) involved in the release of anterior pituitary hormones has been the subject of much investigation, but all that can be said at present is that it is unlikely to be chemically related to adrenaline, histamine or acetylcholine.

Adenohypophysial function is influenced by hormones in the general circulation. Thus pituitary function is depressed by gonadal, thyroid and adrenal cortical hormones, and it is possible that under ordinary conditions regulation of pituitary activity is achieved by this 'feed-back' mechanism which may involve the hypothalamus as an intermediary. There can be no doubt that the secretions of the hypophysis are influenced through the nervous system. Many kinds of mental and physical activity, coitus and suckling for example, influence pituitary activity. It is possible that the nervous and emotional stimuli act via the hypothalamus.

Emotional stress in the rabbit produced, for example, by electric shocks applied to the skin liberates ACTH which acts on the adrenal cortex and causes a lymphopœnia. This effect is abolished by hypophysectomy or by destructive lesions placed in the posterior hypothalamus (mamillary bodies or posterior tuber cinereum). Stimulation of the posterior hypothalamus, but not of the anterior, also causes a lymphopœnia. The pathway involved, hypothalamus → hypophysial portal vessels → anterior pituitary → ACTH release → adrenal cortex → cortical hormone release, appears to be the pathway concerned in quiet conditions for maintaining secretion of ACTH but it is not necessarily the only pathway followed in stress reactions. How the hypothalamus is activated by stress is not clear but the following observations indicate that there may be more than one route. Stimulation of certain cortical areas, for example the orbital surface of the frontal gyrus and the incus, produces eosinopœnia. Stress stimuli as well as activating the posterior hypothalamus increase electrical activity in the anterior nuclei of the thalamus and in the cingulate gyrus.

THE ADENOHYPOPHYSIS

Methods used to determine the function of the adenohypophysis have included removal or implantation of pituitary tissue, the injection of extracts prepared in various ways and, more recently, the administration of more or less pure preparations derived from them. The study of syndromes in man and their correlation with pathological lesions, the site and extent of which was determined after death, has yielded important information. Although at various times it has been claimed that ten to fifteen distinct hormones were produced in the adenohypophysis, to-day only six are believed to have separate identity. They are all proteins and have now been obtained in a highly purified state (Table 53, 8).

Since certain of these hormones produce their main effects indirectly through their actions on other endocrine glands ('target organs') they may be distinguished as *trophic hormones*, and have been given names ending in 'trophic' or 'trophin.' This group includes the pituitary gonadotrophins and the thyrotrophic and adrenocorticotrophic hormones. A second group comprises principles acting directly upon tissues; this group includes the growth hormone and the lactogenic hormone, prolactin.

The Thyrotrophic Hormone (TSH).—Injections of acid or alkaline extracts of the adenohypophysis increase the functional activity of the normal thyroid; conversely ablation of the adenohypophysis results in atrophy of the thyroid gland and a pronounced fall in metabolic rate which can be restored to normal either by injections of pituitary extracts or by transplants. Such extracts initiate metamorphosis in tadpoles and other larvæ just as does thyroid extract. In man hypopituitarism is accompanied by a low basal metabolic rate and even by myxœdema, and conversely in acromegaly where the adenohypophysis is overactive the basal metabolism is frequently above the normal. After thyroidectomy in animals hypophysectomy is without effect on metabolic rate. These effects are due to a thyroid-stimulating hormone (TSH). This is a protein, active in doses of a few micrograms, with a direct action on thyroid tissue since it produces hyperplasia in slices of thyroid in tissue culture and in thyroid transplants. Work with radioactive iodine shows that TSH controls the rate of iodine uptake by thyroid tissue and suggests that it influences the synthesis of thyroxine from diiodotyrosine.

The concentration of TSH in the blood in myxœdema is relatively high while in hyperthyroidism it is low, the hormone having been taken up in large amounts by the hyperactive thyroid tissue. The secretion of TSH by the adenohypophysis is apparently regulated by the level of thyroid hormone in the blood, high concentrations of the latter inhibiting secretion. This phenomenon is probably but one example of a more general effect for the output of gonadotrophins by the adenohypophysis is likewise governed by the blood concentration of ovarian hormones (p. 971).

TABLE 53, 8

Adenohypophysial Protein Hormones

		Mol. wt.	
Growth hormone (somatotrophin)		49,000	Obtained crystalline.
Thyrotrophic hormone (thyrotrophin, thyroid stimulating hormone, TSH)		10,000	
Adrenocorticotrophic hormone (ACTH, adrenotrophin, adrenocorticotrophin)		20,000	Dissociates into several smaller active proteins.
Gonadotrophin	Interstitial cell stimulating hormone (ICSH, luteinizing hormone, LH)	varies	40,000 (sheep); 100,000 (pig).
	Follicle stimulating hormone (FSH)	67,000	
Lactogenic hormone (prolactin, luteotrophin, mammotrophin)		32,000	Obtained crystalline.

It is now generally agreed that the exophthalmos which is so striking a feature in many patients with hyperthyroidism is not due to overactivity of the thyroid *per se*. Even when the thyroid is grossly overactive exophthalmos may be absent and it may appear for the first time after the hyperthyroidism has been relieved by thyroidectomy. Moreover, exophthalmos is not produced in animals by administration of thyroid but it may follow injections of adenohypophysial extracts even in thyroidectomized animals. It is believed therefore that exophthalmos is produced by a principle in the adenohypophysis which may be TSH, although it has been suggested that a separate ophthalmotrophic factor is responsible. There is evidence that the protrusion of the eyeball is produced by an œdema localized to the retro-orbital tissues, for which the hormones of the suprarenal cortex and neurohypophysis have been held responsible since they influence the amount of fluid retained in tissues.

The Adrenocorticotrophic Hormone (ACTH).—Hypophysectomy is invariably followed by considerable atrophy of the adrenal cortex, though not of the medulla, and the daily administration to animals of extracts of adenohypophysis is regularly followed by marked hyperplasia in the adrenal cortex. Clinical studies, also, indicate that in conditions known to be associated with hypofunction of the adenohypophysis (dwarfism and Simmonds' disease) there is marked atrophy of the adrenal cortex and evidence of cortical hypofunction. Contrariwise a secreting tumour in the pituitary (basophilic adenoma) is accompanied by hypertrophy of the adrenal cortex and clinical signs of overactivity.

The physiological properties of ACTH are entirely attributable to stimulation of the adrenal cortex, a subject that has already been referred to in pp. 936–941. Briefly the main effects produced by injection of ACTH in man are

(1) Retention of salt and water and increased elimination of potassium —the aldosterone or mineralo-corticoid effect.
(2) Elevation of blood sugar and glycosuria; increased elimination of nitrogen in urine; depression of circulating eosinophil and lymphocyte concentrations. This is the cortisone or gluco-corticoid effect.
(3) Increased elimination of 17-ketosteroids—the androgenic effect.

ACTH is destroyed by the digestive enzymes and therefore has to be given by injection, either intramuscular or intravenous. The hormone is rapidly inactivated by liver and kidney and its half-life in the blood is only about five minutes. After a single injection of ACTH there is hypertrophy of the zona fasciculata and zona reticularis of the adrenal cortex and at the same time the concentration of lipid and of ascorbic acid falls in all the zones. If injections are continued hypertrophy continues and the lipid content of the cells is restored.

In certain circumstances the relationship between the dose of ACTH and reduction of the ascorbic acid content of the adrenal cortices of hypophysectomized rats is sufficiently close to form the basis of a method of assay of the hormone (Sayers' method). The biological potency of preparations of ACTH is now expressed in international units; one i.u. is defined as the activity of 1 mg. of a sample of material, known as the international standard, kept at the National Institute for Medical Research in London. There is not always a close correlation between the potency of extracts assayed in this way and their clinical effects and some evidence suggests that there may be more than one compound with adrenocorticotrophic activity. Preparations containing as much as 100 to 200 i.u. per mg. have been made.

Maximal stimulation of the adrenal glands in man is achieved by a dose of about 10 units given by intramuscular injections every six hours. The body's response to an injection of ACTH is shown most easily by an increase in the output of 17-ketosteroid in the urine above the normal (p. 953) or a fall of more than 50 per cent. in the eosinophil count in the blood measured four hours after the injection. These measurements are also useful as diagnostic tests of atrophy or destruction of the adrenal cortex (Addison's disease, p. 939) for in such circumstances the cortex cannot be stimulated by ACTH.

ACTH was isolated in 1943 from pituitaries both of the pig and the sheep. It is a simple protein with an isoelectric point about pH 4·6 and a molecular weight about 20,000, but although this material is apparently homogeneous, it has recently been shown to consist of several dissociable components.

Hog ACTH for example yields eight active proteins. One of these, the β-ACTH with a molecular weight of 4,500, has the following sequence of amino-acids (the contractions employed are the same as those used on p. 49 for the structure of insulin):—

Ser-Tyr-Ser-Met-Glu-His-Phe-Arg-Tyr-Gly-Lys-Pro-Val-Gly-Lys-Lys-Arg-Arg-Pro-Val-Lys-Val-Tyr-Pro-Ala-[Gly-Glu-Asp]-Asp-Glu[NH_2]-Leu-Ala-Glu-Ala-Phe-Pro-Leu-Glu-Phe.

GONADOTROPHIC HORMONES.—Removal of the adenohypophysis from immature animals prevents the development of the gonads, which are stimulated by extracts prepared from it. Since the two hormones responsible for these effects have no direct action on the secondary sex organs, but only an indirect one through the ovaries and testes, they have been named gonadotrophic hormones. The follicle stimulating hormone (FSH) is a glycoprotein with a molecular weight of about 67,000 and an isoelectric point about pH 4·5. The constitution of the luteinizing hormone (LH), which is also referred to as the interstitial cell stimulating hormone (ICSH), varies considerably according to the species investigated, but it is also a glycoprotein.

The physiological properties of the gonadotrophic hormones have already been described on pp. 968 to 971.

THE GROWTH HORMONE (somatotrophin).—The growth hormone, obtained relatively free from gonadotrophic, lactogenic, thyrotrophic and adrenocorticotrophic activity, is a heat-labile protein which behaves electrophoretically as a single substance with an isoelectric point at pH 6·85 and a molecular weight of the order of 49,000. It is readily hydrolysed by strong acids and by proteolytic enzymes of the alimentary canal. Daily injections of potent extracts not only produce an increase in size of the injected animal as compared with its litter-mate but also extend the period of growth beyond the normal time, with the result that the injected animal attains a size more than twice its untreated control (Fig. 53, 9). This growth-promoting effect is accompanied by a retention of nitrogen and it has been suggested that the hormone influences growth through its control of protein metabolism. It increases the total oxygen consumption while depressing carbohydrate oxidation and increasing fat metabolism as indicated by a lowering of the R.Q. It must be realized clearly that growth is not dependent on the growth hormone alone; the growth of thyroidectomized rats, for example, is not accelerated by injection of anterior pituitary extracts and these are not effective in accelerating growth in the absence of a supply of insulin from the animal's pancreas.

Hyperglycæmia and glycosuria follow injections of adenohypophysial extracts into animals on normal diets, and in 1937 F. G. Young reported that dogs given daily injections of crude extracts developed a permanent diabetic state, with complete degeneration of pancreatic islet tissue, which persisted after the injections were stopped. The degeneration is apparently due to prolonged stimulation and subsequent exhaustion of the islets by the continuous hyperglycæmia and excessive gluconeogenesis produced by the pituitary extract. It has recently been shown that intact adult cats can be made diabetic by repeated injections of pure growth hormone. However, since diabetes can also be produced by treatment with ACTH, it may be that the so-called diabetogenic effect of adenohypophysial extracts is the result of the presence in them of both hormones. Animals treated with extracts made from the adenohypophysis are rendered resistant to the action of insulin.

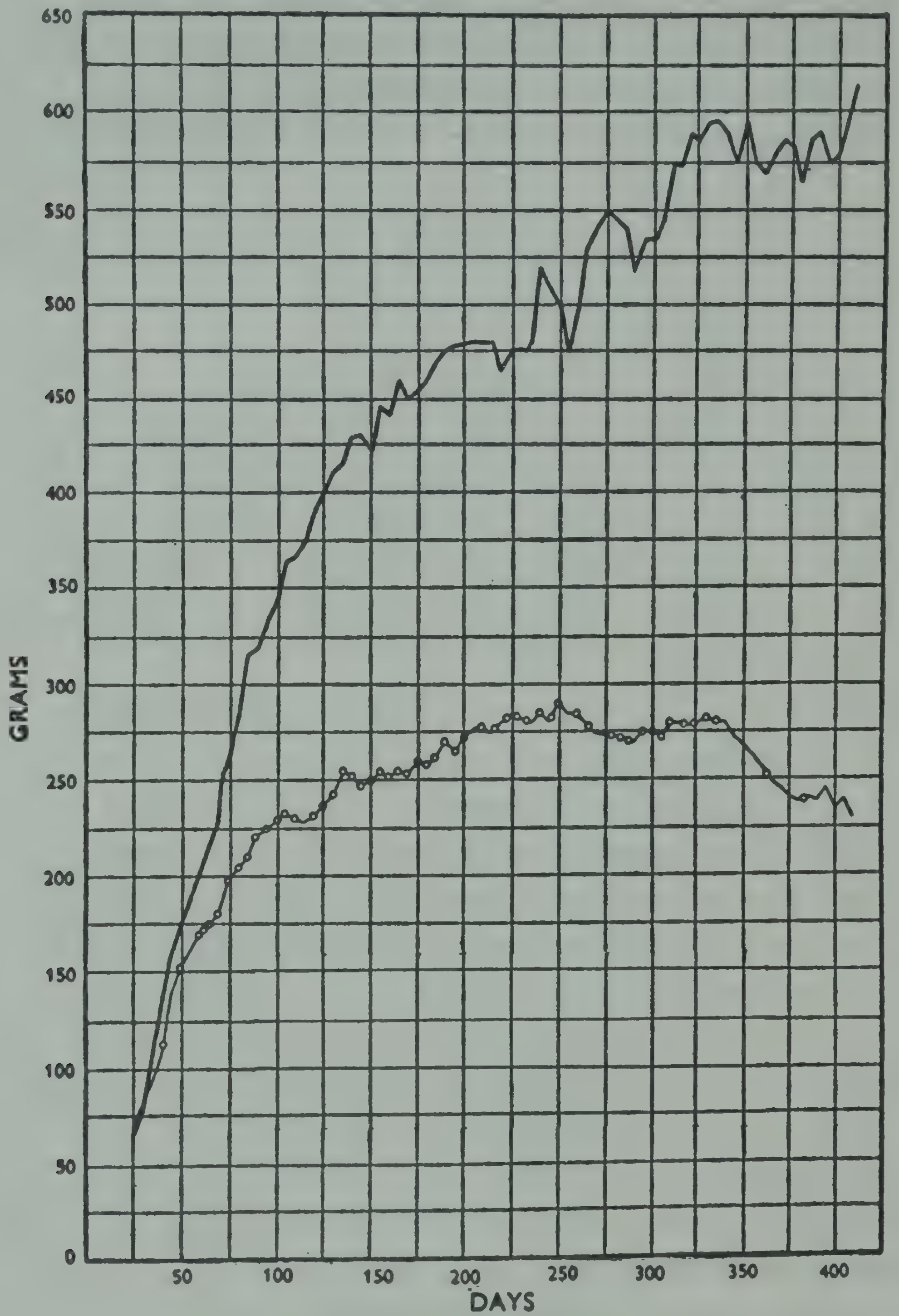


FIG. 53, 9

The effect of an extract of the adenohypophysis on growth. The upper line is the growth curve of a female rat treated for 400 days, the lower line that of an untreated litter-mate. (Redrawn from H. M. Evans (1923–24), *Harvey Lect.* **19**, p. 212. Philadelphia: Lippincott.)

THE LACTOGENIC HORMONE (prolactin, also called mammotrophin or luteotrophin).—The production of milk depends on the pituitary for it is at once arrested by hypophysectomy, though it may be maintained after the operation by pituitary extracts. After injection of this hormone into male or female pigeons there is an increase in the production of a milk-like fluid, 'pigeon's milk', in the crop gland, and this observation has been made the basis of an assay. Lactogenic hormone has been detected in the pituitaries of all vertebrates so far examined and is present in approximately equal amounts in male and female. Its actions are described in more detail on p. 993.

Prolactin has been isolated from the pituitary glands of sheep and cattle as a homogeneous protein which has been crystallized. Its isoelectric point is pH 5·7 and its molecular weight about 32,000.

THE MELANOPHORE-EXPANDING HORMONE.—Hogben and Winton first showed that extracts made from the posterior part of the pituitary caused darkening of the skin of frogs and that after hypophysectomy the frogs became light in colour. These changes are due to the fact that the skin of the frog, like that of very many cold-blooded animals, contains large cells with branching processes containing granules of melanin (melanophores). When the melanin is aggregated in the body of the cell the animal appears pale in colour but when the pigment is dispersed throughout the cell body and its process the skin is dark. The distribution of pigment in the skin, and therefore the blending of the animal with its environment, is under hormonal control, adrenaline producing blanching and the pituitary hormone darkening. These experiments are important since they provide clear evidence that a hormone can govern the distribution of materials within the cell. Though it is found in extracts of the neurohypophysis this hormone is elaborated in the pars intermedia of the adenohypophysis of animals. Many preparations of ACTH contain much melanophore-expanding activity but the material is believed to be distinct from the adrenocorticotrophic substance.

CLINICAL DISORDERS OF ADENOHYPOPHYSIAL FUNCTION

As might be expected from the foregoing account of the physiological properties of adenohypophysial extracts the effects of disease of the adenohypophysis are very variable and involve many organs. Although the early stages of pituitary dysfunction are not easy to recognize and the later stages complex and difficult to analyse it is possible to define a number of fairly clear-cut disorders.

GIGANTISM AND ACROMEGALY.—These are produced by the long-continued hypersecretion of the adenohypophysis and are mainly, though not necessarily wholly, the result of overproduction of the growth hormone by acidophil cells. The conditions are similar in that the tissue overgrowth (hypertrophy) affects in varying degrees all the tissues of the body, but they differ in that, while in the former the excessive growth begins before the epiphyses close, in the latter the condition begins in adult life.

In gigantism the abnormal growth is as a rule well advanced before the age of puberty, with the result that the patient may attain a height of seven to eight and a half feet. The growth of the limbs is relatively greater than that of the trunk and the tallness is largely the result of excessive growth of the legs. The internal organs are also hypertrophied and the muscles, at least in the early stages, may be very well developed. Such individuals are usually mentally subnormal and die before the age of twenty.

In acromegaly (Fig. 53, 10) the disturbance of growth leads to bony thickening and deformity rather than to increased stature but in this condition, as in gigantism, the internal organs may be enlarged. The most striking clinical feature of acromegaly is, however, the generalized coarsening of the features due to thickening of the skin and subcutaneous tissues, and the progressive enlargement of the head, the hands and the feet. Increased sexual function with mammary hyperplasia and lactation are occasionally found. Elevated B.M.R. and disturbance of carbohydrate metabolism, as evidenced by hyperglycæmia and glycosuria, are frequently encountered. Since the excessive secretion from the adenohypophysis is often associated with a tumour consisting of acidophil cells the acromegalic patient may have symptoms referable to pressure upon adjacent parts.

Cushing's syndrome is a complex phenomenon characterized by obesity, redness of the skin, increased growth of hair, high blood-pressure, osteoporosis, and hyperglycæmia with glycosuria. These effects are produced by overactivity of the adrenal cortex, the cause of which is often obscure, or by a tumour of the cortex. However, Cushing's syndrome can be associated with tumours of the basophilic cells of the adenohypophysis which are presumably producing too much ACTH. A similar syndrome has been found with tumours of the thymus and of the ovary. Because of the close functional connexion between the hypothalamus and the adenohypophysis Cushing's syndrome could, theoretically at least, be produced by disturbance in the hypothalamus.

Insufficiency of the secretions of the adenohypophysis leads to recognizable syndromes which are in direct contrast to those mentioned above as being due to overactivity. Destructive lesions of the anterior part of the pituitary are, as might be expected, followed by greater or less degrees of atrophy in other endocrine organs, so that the great variety of signs and symptoms seen are in many cases not primarily due to the disorder of the pituitary but to the defective functioning of other endocrine glands.

PITUITARY DWARFISM (ATELEIOSIS).—Insufficiency of the adenohypophysial secretions in childhood results in delayed growth and a form of dwarfism, with permanently retarded sexual development. Pituitary dwarfs are usually well-proportioned, alert and intelligent, and able to earn their own living on the stage or in the circus, but occasionally other symptoms of hypopituitarism may be found such as are described below. The insufficiency of the adenohypophysial secretion which is responsible for the delay in growth is usually the result of compression of the anterior part of the gland by a slowly growing cystic tumour derived from the epithelium of the buccopharyngeal pouch (p. 998).

PANHYPOPITUITARISM (SIMMONDS' DISEASE).—This is the result of adenohypophysial insufficiency in the adult. It is usually due to destruction of the adenohypophysis following an interruption of its circulation at parturition. There is genital atrophy, gradual regression of secondary sex characteristics, loss of body hair and occasionally a striking appearance of premature senility and sometimes loss of weight. The basal metabolism is reduced, the body temperature subnormal, the heart rate low and the arterial pressure below the normal level. Carbohydrate metabolism is grossly disturbed and patients develop severe and sometimes fatal hypoglycæmia. Treatment with extracts of the adenohypophysis may relieve many of these features but it rarely succeeds in restoring the patient to normal health.

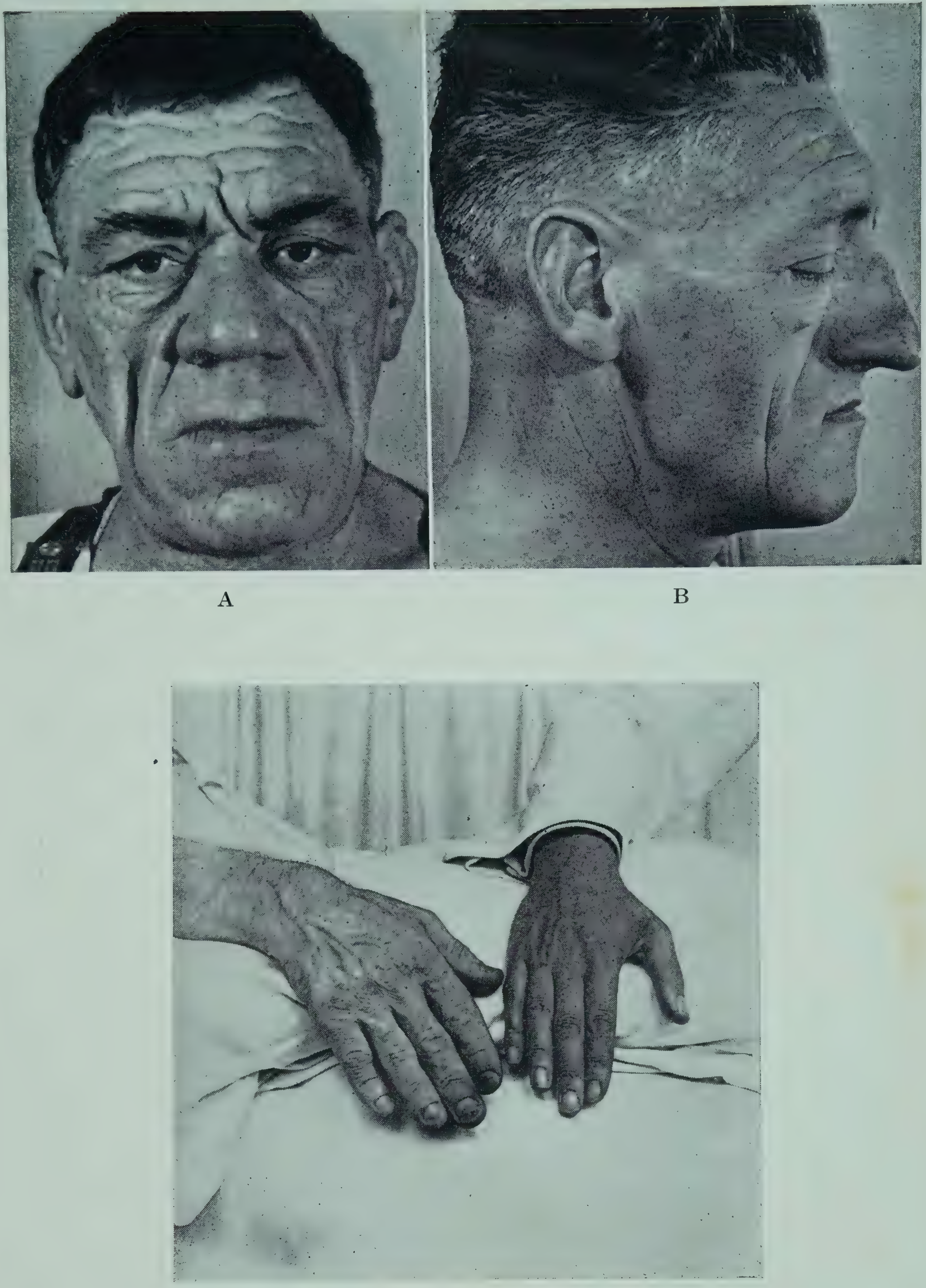


Fig. 53, 10

A. Front view of the face of a patient with acromegaly to illustrate the gross thickening of the soft tissues. (*By courtesy of Raymond Greene and the "Practitioner".*)

B. Profile of the face of another patient to show early changes in the bones of the face. Note the massive supra-orbital ridges and the wide angle of the jaw.

C. The hand of an acromegalic patient contrasted with the hand of a normal man.

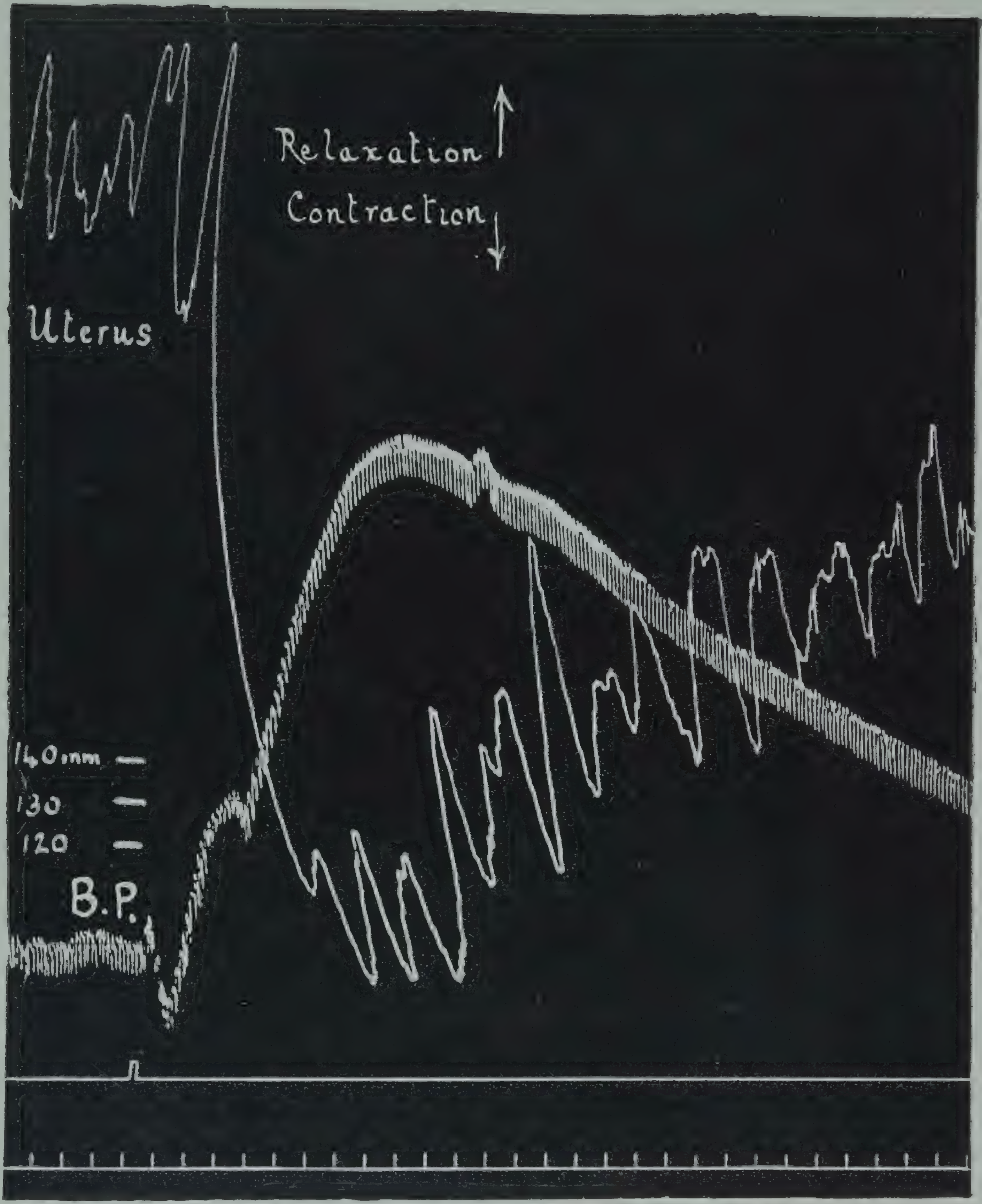


FIG. 53, 11

The oxytocic action of pituitary extract. The original figure of the effect by H. H. Dale (1906), *J. Physiol.* **34,** 195. The sympathetic motor effects had been abolished by a previous dose of ergot. At the signal, extract from 0·4 g. dried ox-pituitary was given intravenously, producing the rise of blood pressure and the contraction of the uterus. Pregnant cat. Time in 10 sec.

THE NEUROHYPOPHYSIS

The injection of an aqueous extract of the pituitary gland into animals was shown by Oliver and Schäfer in 1895 to produce a marked increase of arterial pressure which lasted about ten times longer than that following an injection of adrenaline giving a similar rise of pressure. Since that time pituitary extracts have been shown to have effects on many tissues but only three of them are sufficiently different to justify the tentative conclusion that the neurohypophysis elaborates more than one hormone.

Extracts of the neurohypophysis have, then, three main effects. They bring about contraction of the uterus (oxytocic effect); they increase arterial pressure (vasopressor effect); they reduce the output of urine (antidiuretic effect). By repeated fractionation oxytocic extracts with little pressor effect have been obtained and also extracts with pronounced pressor and antidiuretic effects but with little oxytocic activity. Simple acid extracts of the neurohypophysis stimulate plain muscle directly but the action, unlike that of adrenaline and noradrenaline, is exerted widely on smooth muscle regardless of whether it is normally stimulated by sympathetic or parasympathetic nerves. Such extracts therefore produce contraction of the smooth muscle of the blood vessels, uterus, intestine, ureters, gall-bladder and bile ducts. In the lactating female extracts of the neurohypophysis lead to expression of milk by contraction of the muscle of the mammary ducts and alveoli.

It has been suggested that the activity of the neurohypophysis could be attributed to a protein of molecular weight between 20,000 and 30,000 upon which the physiologically active principles are adsorbed. However that may be, two active fractions, *oxytocin* and *vasopressin* have been identified and synthesized by Du Vigneaud. Both are octapeptides of molecular weights 1,025 and 1,102 respectively.

Oxytocin.—The most convincing demonstration of the action of the oxytocic principle is that electrical stimulation of the supraoptico-hypophysial tract produces contractions of the uterus (provided that the animal has been treated with œstrogens) even when the connexions between the head and body are confined to the carotid arteries and jugular veins. Lesions of the tract, experimentally produced, interfere with parturition in both cats and guinea-pigs. It is probable that oxytocin has an important function in parturition, at any rate in some animals, but its precise role is not yet known. Oxytocic extracts with only a small pressor effect find a ready application in obstetrics where, after a hæmorrhage from the uterus at parturition, it is desirable to produce a firm and sustained contraction of the uterine muscle without elevating the blood pressure.

Vasopressin.—The pressor fraction of the neurohypophysis is a powerful vasoconstrictor substance exerting its action on the minute vessels as well as upon arterioles, so that marked pallor of the skin is one of its effects. The increase of arterial pressure is followed by slowing of the heart and passive dilatation of the cerebral and renal vessels, which are apparently not constricted by vasopressin. Repeated injections of pressor extracts produce depression of the blood pressure, an effect which is said to be due to constriction of the coronary vessels and interference with cardiac contractions rather than to a loss or reversal of the action upon arterioles. Removal of the neurohypophysis from animals, however, is not normally followed by a fall in pressure and there is little evidence that vasopressin is involved in the physiological control of blood pressure.

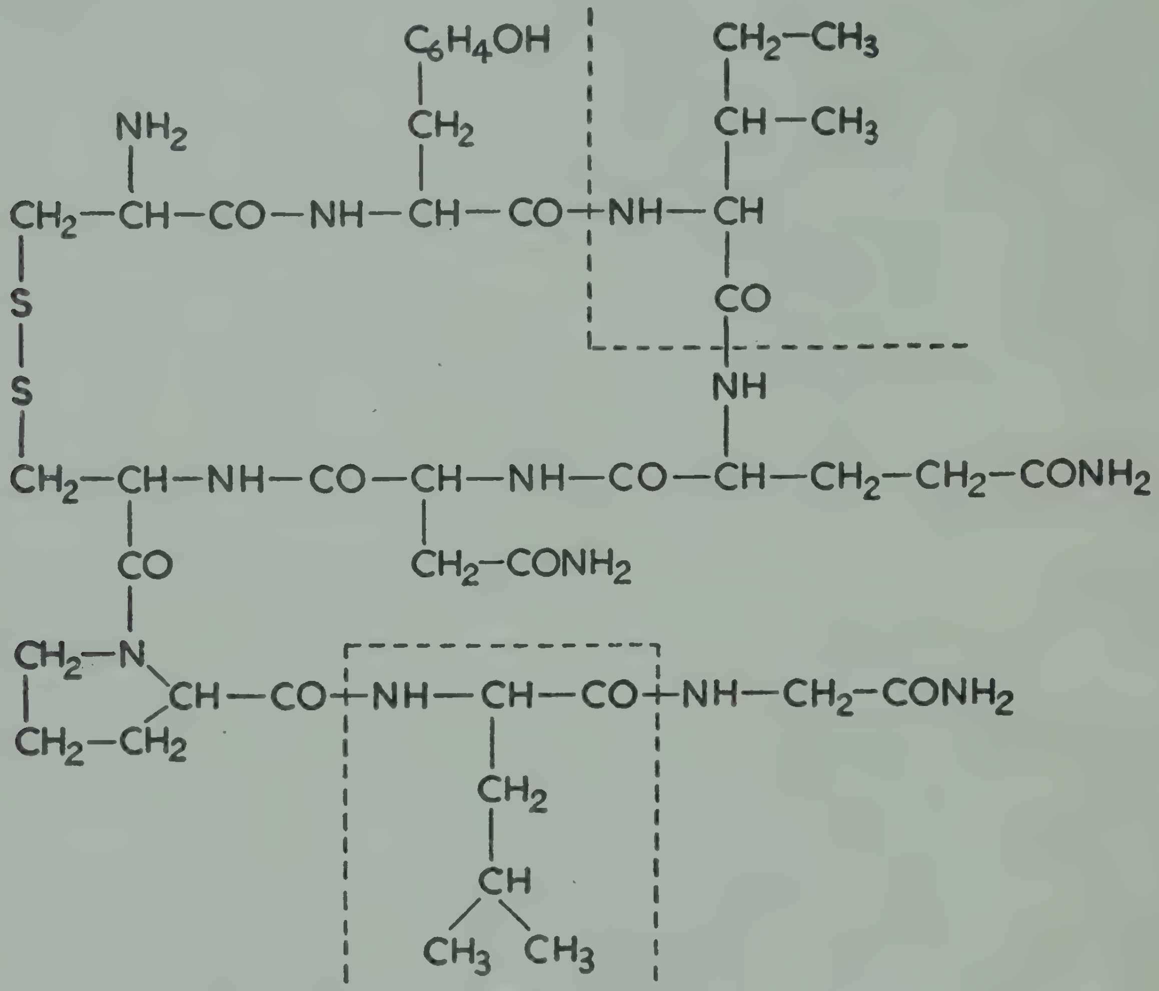
The sequence of amino-acids in bovine and hog oxytocin is:—

Cy.S-Tyr-Isoleu-Glu-Asp-Cy.S-Pro-Leu-Glycinamide

and in bovine vasopressin is:—

Cy.S-Tyr-Phe-Glu-Asp-Cy.S-Pro-Arg-Glycinamide

where the contractions used on p. 49 are again employed. The full structure of oxytocin is therefore as follows:—



In vasopressin of bovine origin the structure is very similar to that of oxytocin, the only differences being the replacement of isoleucine (top right hand) by phenylalanine and of leucine (bottom centre) by arginine. Vasopressin from hog glands contains lysine in place of arginine.

Antidiuretic Principle.—The pressor principle of the adenohypophysis, when administered to a mammal in much smaller doses than are required to raise the blood pressure, produces a marked fall in the rate of urine flow (oliguria). Conversely experimental removal of the neurohypophysis, or division of the supraoptico-hypophysial tract, is followed by increased elimination of water (polyuria) although not, as we have seen, by a fall in blood pressure. For these reasons vasopressin is usually referred to as ADH (antidiuretic hormone). It is essential to the maintenance of fluid equilibrium and acts directly on the renal tubules to augment the re-absorption of water from the glomerular filtrate (see page 613). The state of dehydration, in which fluid must be conserved in the body, leads to an appreciable increase in the amount of antidiuretic activity in the infundibular process. At the same time there is an increase of antidiuretic potency in the blood and urine. The œdema of cardiac failure and of malnutrition is associated with an increase of antidiuretic potency in the urine and this is augmented also in chronic liver disease with accumulation of fluid in the peritoneal cavity (ascites).

If ADH is injected into an animal from which the liver and kidneys have previously been removed the hormone disappears from the blood much more slowly than it does in the intact animal. Furthermore, large doses produce a greater rise of arterial pressure when injected into the femoral vein of an animal than when given into the splenic vein. These experiments suggest that ADH is destroyed, or removed from the circulation, by the liver.

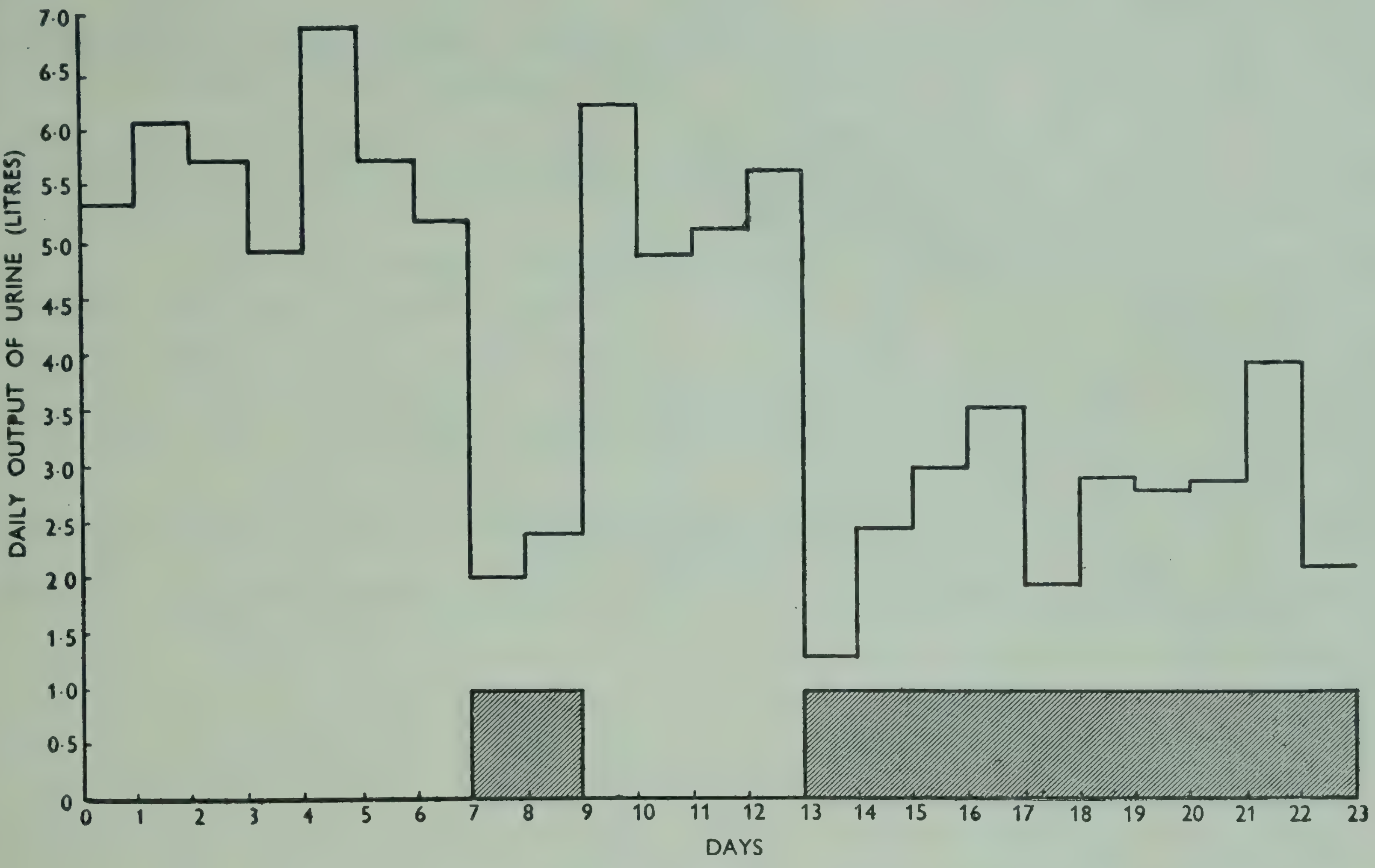


FIG. 53, 12

The effect of extracts of the neurohypophysis on the daily output of urine of a male patient, aged 24, with diabetes insipidus. During the first shaded area three subcutaneous injections a day of a simple extract were given on each of two successive days. During the second shaded area the patient was treated with daily injections of the same extract.

If the infundibular process of the neurohypophysis is removed or destroyed by disease the condition known as *diabetes insipidus* arises, in which as much as thirty litres of urine a day may be passed. The urine in these cases is pale with a specific gravity little more than unity and a low chloride content. Injection of small amounts of active extracts can reduce the elimination of water to normal (Fig. 53, 12). From Fig. 34, 17 it is clear that diabetes insipidus can be produced by lesions of the hypothalamus as well as by injuries to the region of the tuber cinereum.

REFERENCES

CHOW, B. F. (1944). The purification and properties of certain hormones, in *Advances in* **1,** *Protein Chemistry*, 153–185. New York: Academic Press Inc.

Ciba Foundation Colloquia on Endocrinology, vols. I to V. London: Churchill.

CROOKE, A. C. (1955). Tests of adenohypophyseal function. *Lancet*, ii, 1045–1050.

EVANS, H. M., MEYER, K. & SIMPSON, M. E. (1933). *Growth and Gonad-stimulating Hormone of the Anterior Hypophysis.* Berkeley: Univ. of Calif. Press.

FRIEDGOOD, H. B. (1945). *The Endocrine Function of the Hypophysis.* Oxford University Press.

GREENE, R. (Ed.) (1948). *The Practice of Endocrinology.* London: Eyre & Spottiswoode.

HARRIS, G. W. (1955). Neural control of the pituitary gland. *Physiol. Soc. Monogr. No.* 3. London: Arnold.

LI, C. H. (1949). The chemistry of gonadotrophic hormones. *Vitam. & Horm.* **7,** 223–252.

LI, C. H. & EVANS, H. M. (1947). The properties of the growth and adrenocorticotrophic hormones. *Vitam. & Horm.* **5,** 197–231.

MARTIN, L. & HYNES, M. (1954). *Clinical Endocrinology.* London: Churchill.

PINCUS, G. (Ed.) (1948–50). *Recent Progress in Hormone Research.* New York: Academic Press Inc.

PINCUS, G. & THIMANN, K. V. (1948). *The Hormones.* New York: Academic Press Inc.

SMART, G. A. (1954). Pituitary function. *Brit. Med. J.* i, 1086-1089.

VAN DYKE, H. B. *The Physiology and Pharmacology of the Pituitary Body*, vol. i, 1936, vol. ii, 1939. Chicago: University Press.

VERNEY, E. B. (1947). The antidiuretic hormone and the factors which determine its release. Croonian Lecture. *Proc. Roy. Soc. B*, **135,** 35–105.

WHITE, A. (1946). Preparation and chemistry of anterior pituitary hormones. *Physiol. Rev.* **26,** 574–608.

CHAPTER 54

CELL DIVISION AND HEREDITY

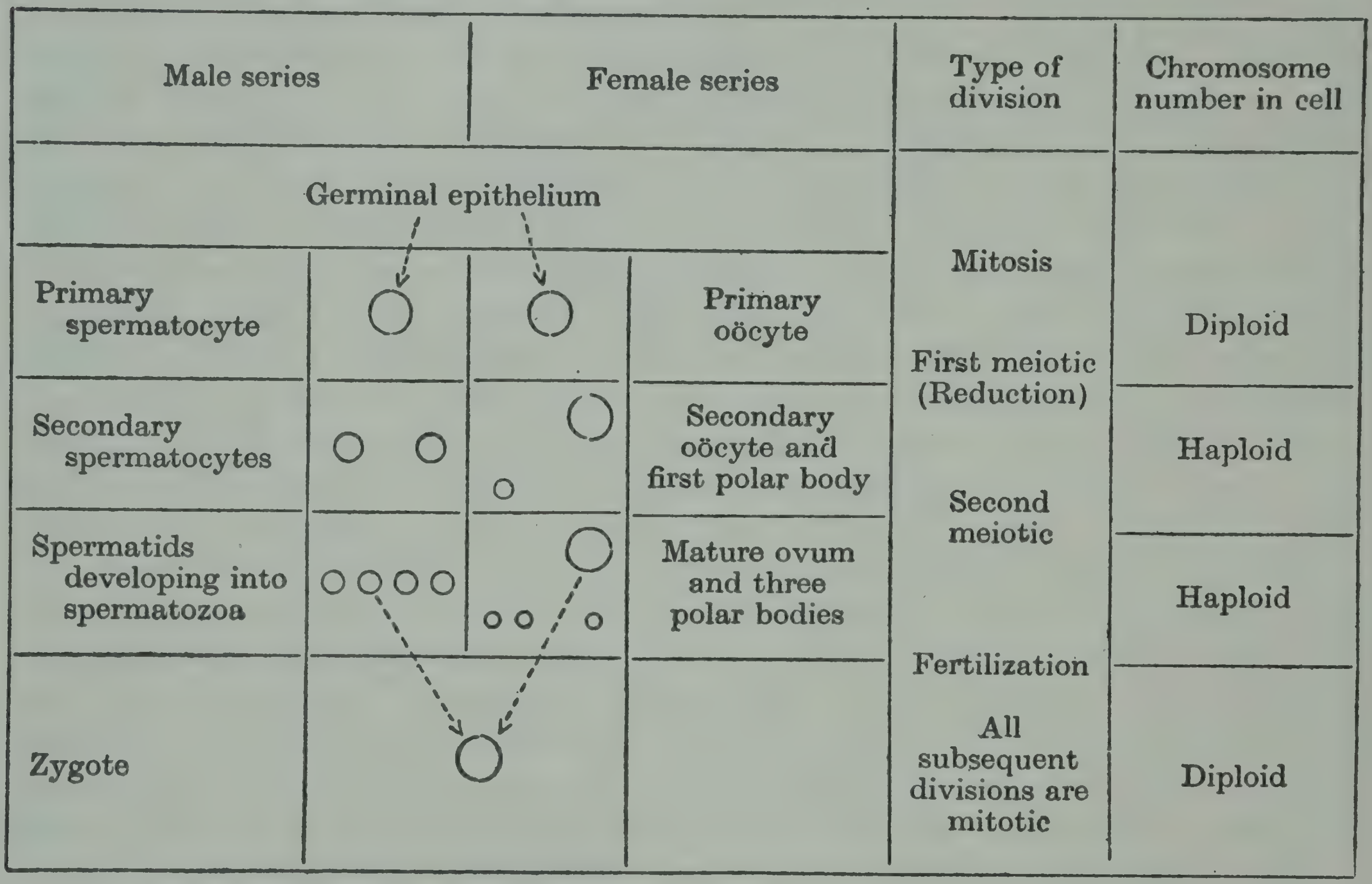
We have seen how cells are adapted to the specific function which they have to perform in the different tissues and that they are therefore often highly differentiated morphologically. This differentiation begins during the embryonic period and continues during growth and as long as the cells have the ability to multiply. In some tissues, such as epidermis and bone marrow, certain cells remain undifferentiated and these can multiply readily. In glandular tissues, such as liver and kidney, all the cells normally are differentiated but they still have a limited power of regeneration should the organ suffer any damage. The ability to replace any worn out or damaged cell has disappeared in nerve cells which in the adult have lost the capacity to divide. During differentiation the morphological changes which occur in the cell are less marked in the nucleus than in the cytoplasm. Since the nucleus contains the units of heredity it is important that any new cells which are formed should have the same chromosome complement as the parent cell. This orderly division of nuclear material is known as *mitosis.* At the outset of a division the chromatin becomes more conspicuous by the increase in the amount of nucleic acid (p. 63) and it is soon apparent that each chromosome has reproduced itself and is already 'split.' The nuclear membrane then disappears, a spindle of hyaline fibres is formed between the two centrosomes and the 'split' chromosomes line up on the equator. The daughter halves then separate and travel towards the poles of the spindle so that each new nucleus contains an exact replica of the parent nucleus. The cytoplasm then divides and two new daughter cells are formed. In unicellular organisms mitotic division of the cell results in the formation of two new individuals, and this is actually a special form of reproduction known as binary fission.

The chromosome complement of a multicellular individual is derived equally from its two parents. Its cells contain two numerically equal chromosome sets or genomes and the individual is said to be diploid. The recognizably different chromosomes of one genome can usually be paired up with those of the other genome and these 'like' chromosomes of maternal and paternal origin are said to be homologous. In man there are 48 chromosomes, 24 of which are derived from each parent.

The process of sexual reproduction involves the conjugation of two cells called *gametes* with union of their nuclei to form a *zygote* or fertilized ovum. The gametes arise within gonads from cells which remain in a relatively undifferentiated state until the onset of the reproductive period. These primordial germ cells contain the diploid number of chromosomes and if the gametes derived from them were to retain this full somatic number of chromosomes the number in the zygote would be doubled. Such an increase is prevented by the occurrence of *meiosis*, in which there are two consecutive divisions of the cell but only one division of the chromosomes, during the maturation of the gametes. In this process the chromatin of the nucleus resolves itself into chromosomes as before, but one out of each pair of homologous chromosomes goes to one daughter cell while the second passes to the other daughter cell. This reduction division is immediately followed by a second division in which the chromosomes 'split' longitudinally, the daughter cells receiving the chromosomes in the way characteristic of mitosis. Thus each female gamete or ovum and each male gamete or spermatozoon contains only the haploid number of chromosomes (in man, 24). It should be noted that in ordinary mitosis the diploid chromosome complement is *retained* by the daughter cells but in meiosis the diploid number is *shared* between the daughter cells. The random distribution of the homologous chromosomes may result in a shuffling of the chromosomes of the two genomes so that different combinations of maternal and paternal chromosomes arise in the germ cells produced.

TABLE 54, 1

The Changes in the Cells involved in the Reproductive Process



Male series			Female series	Type of division	Chromosome number in cell
Germinal epithelium				Mitosis	
Primary spermatocyte			Primary oöcyte	First meiotic (Reduction)	Diploid
Secondary spermatocytes			Secondary oöcyte and first polar body	Second meiotic	Haploid
Spermatids developing into spermatozoa			Mature ovum and three polar bodies	Fertilization	Haploid
Zygote				All subsequent divisions are mitotic	Diploid

Both ova and spermatozoa are derived from the germinal epithelium. In the seminiferous tubules (p. 948) each primary spermatocyte undergoes meiosis, the first reduction division giving two secondary spermatocytes which on their second division give four spermatids all of the same size (Table 54, 1). In the transformation of a spermatid into a spermatozoon the nucleus is retained in the head and the centrosomes in the middle piece, and the much reduced cytoplasm forms a thin investing layer from which the axial filament of the tail projects.

During maturation of the ovum analogous changes occur, but the cytoplasm is not divided equally between the daughter cells (Table 54, 1). The ovum (or primary oöcyte) undergoes a reduction division so that each of the daughter cells has the haploid number of chromosomes; one of the two daughter cells, the secondary oöcyte, retains most of the cytoplasm; the other smaller cell is cast off as the *first polar body.* The secondary oöcyte now divides again and throws off another polar body; the first polar body may or may not divide again. This process of maturation of the ovum probably begins in man just before ovulation.

In most mammals, including the human species, fertilization occurs in the uterine tube by the entry of a spermatozoon into a mature ovum, the zygote formed by this union having the diploid number of chromosomes. Thus a new individual begins life with a chromosome number derived equally from its parents and this number is usually retained in all its cells since all subsequent divisions are mitotic.

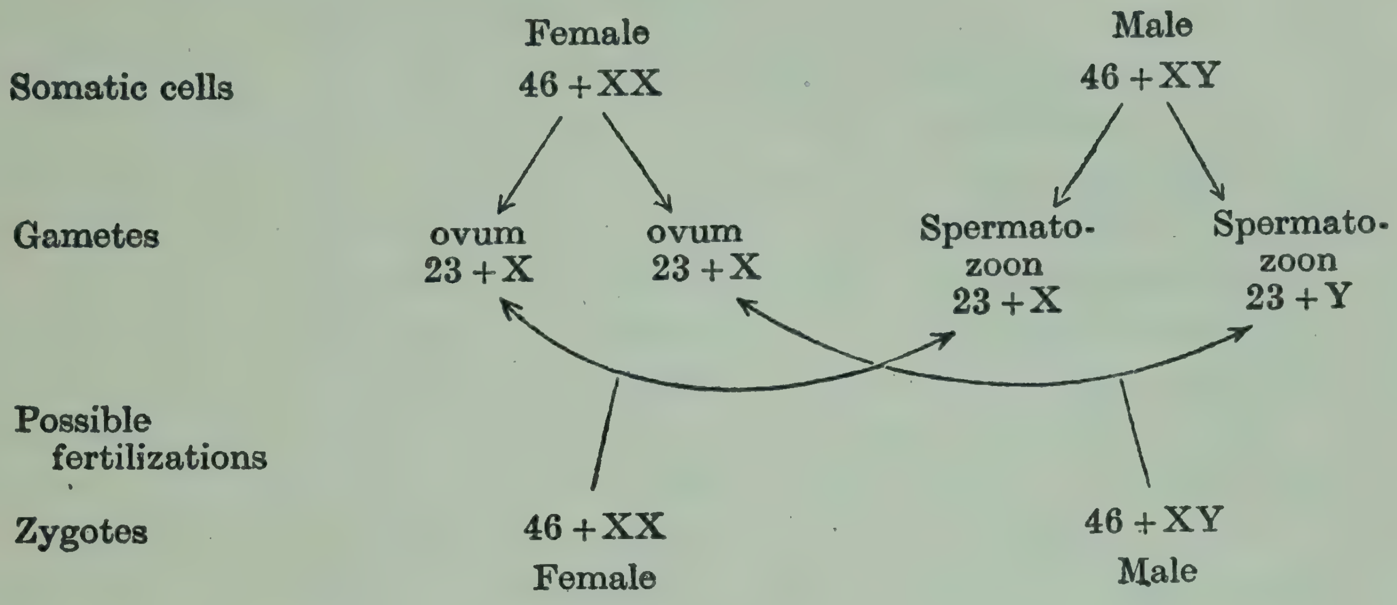


FIG. 54, 2

This diagram shows that the sex of the future individual is determined by the kind of spermatozoon which fertilizes the ovum.

In many animals the members of one pair of chromosomes differ in the two sexes and they are, therefore, referred to as the sex chromosomes. In the mammalian female the two components of the pair are similar and may be denoted by XX; since at meiosis the chromosome material is shared equally by the daughter cells all the ova contain a single X chromosome. In the male, on the other hand, the pair consists of one X chromosome and a much smaller sex chromosome, Y; the pair of male sex chromosomes is, therefore, XY. At the reduction division X passes into one secondary spermatocyte and Y into the other. The subsequent meiotic division results in two X spermatids and two Y spermatids which ultimately develop into two X spermatozoa and two Y spermatozoa. It follows from this that the sex chromosome pattern in the zygote is of two kinds, XX or XY, according to whether the mature ovum, X, has been fertilized by an X spermatozoon or a Y spermatozoon. The XX zygote develops into a female and the XY zygote into a male (Fig. 54, 2).

SEX RATIO

Since in the course of spermatogenesis X and Y spermatozoa are produced in equal numbers, a 1:1 sex ratio in the fertilized eggs would be expected. In other words, in a large group of offspring there should be an equal number of males and females. Actually the *primary sex ratio*, that is the ratio between males and females at conception (fertilization), considerably exceeds unity as mentioned below but there is no satisfactory explanation for this. It might be due to the ova having a slight ability to select the Y spermatozoa in preference to the X spermatozoa but this is unlikely. It might be that conditions in the genital tract, perhaps pH or temperature, favour the Y spermatozoa, but various attempts to control the sex of the offspring, for example by altering the pH of the vagina, have met with no success. Control of sex is unlikely to be achieved until some method is found of separating the X spermatozoa from the Y spermatozoa.

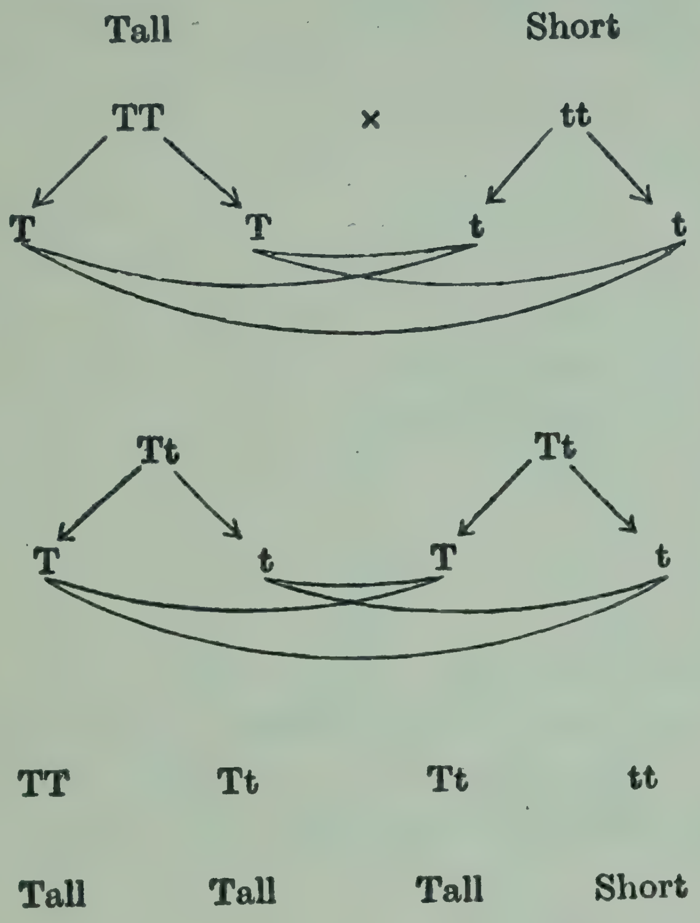
In Great Britain, in 1947, for every 100 girls born there were 105 boys, giving a *secondary sex ratio* of 105 to 100. Approximately 130 male fœtuses perish during the last three months of intrauterine life for every 100 female fœtal deaths and the proportion of male stillbirths is even higher in the third and fourth months of pregnancy. The primary sex ratio must, therefore, greatly exceed 105 to 100 to allow for the wastage of males; one estimate puts it as high as 108 to 100. Since the death rate in males is higher we may conclude that they are more fragile than the females. If the conditions which attend pregnancy are unfavourable to the fœtus, because of disease or malnutrition for example, there may be more fœtal deaths than normal and it is to be expected that the higher fœtal mortality would affect the males particularly; thus the secondary sex ratio would be relatively low. The sex ratio is thus a kind of biological yardstick by which, for example, conditions in different countries can be compared. An improvement in social conditions is usually associated with a rise in the secondary sex ratio, that is a relative increase in the number of males born.

The selective elimination of the male occurs after birth as well as before and has nothing to do with the greater occupational hazards to which he is exposed; it is probably related in some way to his higher metabolic rate. In the group aged 20 to 24 years (in 1947) there was a small excess of males over females in the population of the United Kingdom; between the ages of 25 and 29 the females began to outnumber the males, until in the group aged 85 and over there were more than twice (2·2 times) as many women as men. To be born is a more dangerous adventure for the male than the female and, since there is a higher selective mortality after birth as well as before, the expectation of life is greater at all ages in the female.

PHENOMENA OF HEREDITY

It is now agreed that the *genes* which carry heritable characters are located on the chromosomes and are arranged in an orderly manner along the chromosome fibre which is thus a 'gene string.' Each gene for any particular characteristic has its counterpart in the corresponding locus on the homologous chromosome and these two genes form an *allelic pair*. When both members of a pair of genes carry the same characteristic, say tallness, then the individual is said to be *homozygous* with respect to that character; when one of a pair carries tallness and the other shortness the individual is *heterozygous*. The matter may be made plain by considering an actual example originally described by Mendel, who crossed tall (TT) pea-plants with dwarf (tt) ones (Fig. 54, 3). The first generation were all tall plants. Since they all possessed the Tt pair of allelic genes T must be *dominant* over t, and t must be *recessive*. By selfing, that is by self-fertilizing, these Tt plants he obtained one-quarter dwarfs and three-quarters tall plants. If the dwarfs were selfed they bred true; their gene constitution must, therefore, be tt. Only one-third of the tall plants of this generation bred true; they were TT. The remainder were tall hybrids Tt (heterozygous). In man the character of height is controlled by not one but many genes, so that the single symbol T employed above is not here applicable. The symbol T, however, could be employed to represent the ability to taste phenylthiourea in low concentrations. 'Tasters' have one or two dominant 'T' genes and 'non-tasters' have two recessive 't' genes.

At meiosis the segregation of the homologous chromosomes brings about segregation of the allelic genes. It was stated above that during meiosis there was random separation of the members of the homologous pairs of chromosomes; thus complete gene strings might pass unaltered into the resulting nuclei. But during the preparatory phases in the nucleus prior to the reduction division, that is, during pairing of the homologous chromosomes, it is possible for interchange of allelic genes to take place by 'crossing over' of either large or small portions of the chromosomes, involving one or more genes. These pairs of genes control heritable characters and the free assortment of their alleles is thus provided for.



somatic and primordial germ cells

gametes by reduction division

possible modes of fertilization

somatic and primordial germ cells of next generation

gametes by reduction division

possible modes of fertilization

somatic and primordial germ cells of next generation (genotype)

(phenotype)

FIG. 54, 3

As each individual plant or animal possesses in its chromosomes a large number of genes which are assorted during development of the germ cells, it follows that the zygotes produced by their union must vary in constitution. Any two children in a family are not absolutely alike unless, as occasionally happens, they arise from one fertilized ovum. Such individuals are referred to as monozygotic or identical twins.

INHERITANCE OF BLOOD GROUPS

The blood of all human beings can be classified readily into four major hereditary groups and this provides one of the clearest examples of Mendelian inheritance in man. In the production of the four groups, A, B, AB and O three alternative allelic genes are involved and these may be designated G^A, G^B and g. They are responsible for the presence in the red cells of certain agglutinogens and in the serum for certain agglutinins. The three controlling genes lie at one locus on a certain chromosome and, generally speaking, every one of us inherits two of these alternative genes, a single one from each parent. A person of blood group A may have the genic constitution G^AG^A or G^Ag, where G is dominant to g. A person of genotype G^BG^B or G^Bg belongs to group B; in a group AB person with G^AG^B the genes G^A and G^B are of equal strength and work independently. A person who is homozygous for the recessive gene g (genotype gg) can only belong to blood group O. The ABO blood groups (phenotypes) and the corresponding genic constitution (genotypes) are given in the following table:

Phenotype	Genotype
Group A	G^AG^A and G^Ag
Group B	G^BG^B and G^Bg
Group AB	G^AG^B
Group O	gg

With this information it is possible to work out the inheritance of blood groups and to say which groups can occur in the offspring and which are impossible. If different persons of phenotype A are mated the possible phenotypes and genotypes of the offspring may be expressed diagrammatically thus:

Matings . . .	1				2				3			
Phenotypes . . .	A			A	A			A	A			A
Genotypes . . .	G^AG^A			G^AG^A	G^AG^A			G^Ag	G^Ag			G^Ag
Gametes . . .	G^A	G^A	G^A	G^A	G^A	G^A	G^A	g	G^A	g	G^A	g
Possible fertilizations .												
Zygotes . . .	G^AG^A	G^AG^A	G^AG^A	G^AG^A	G^AG^A	G^Ag	G^AG^A	G^Ag	G^AG^A	G^Ag	gG^A	gg
Phenotypes . . .	A	A	A	A	A	A	A	A	A	A	A	O

According to the genotype of the parents the children could have blood group A or O but never group B or AB. The results of other matings can be worked out in a similar way. Actually the inheritance of the A antigen is more complex than this for the A phenotype is now separable into A_1 and A_2, and recent work has proved that there are many sets of blood group alleles, examples of which are the MN.S groups, the Rhesus group and at least six others. Apart from the ABO and Rhesus groups which are important in blood transfusion and in hæmolytic disease of the newborn (p. 382) a knowledge of their inheritance is of relatively little importance for a medical student except in the study of human genetics and ethnology. Blood group inheritance has some forensic importance because, if the blood groups of a mother and her child are known together with that of the putative father, it can be said that the child either could, or could not, be the offspring of parents with their blood groups. It must be realized that this test is useful in a negative way only and that it is not possible to say, even when there is no incompatibility in the blood group inheritance, that a child is in fact the offspring of any particular pair of individuals. With extension of our knowledge it should soon be possible by examination of the blood groups of the mother and child, to determine those of the male parent so accurately that such could be possessed by only one man in a hundred.

Sex-linked Inheritance

Certain constitutional defects, such as colour blindness or hæmophilia, are more common in males than in females. The explanation lies in the fact that those characters are recessive and the harmful genes for them are carried by one of the X chromosomes, say X′. If the other X chromosome does not carry the harmful gene then a female with the constitution 46 + XX′ is not affected, but a male with constitution 46 + X′Y is affected. In a female with XX′ the allelic gene present in the X can mask the recessive gene carrying the abnormality in the X′—in other words the recessive character does not appear in the presence of the dominant normal gene. An XX′ woman married to a normal XY man could have normal sons, XY, and hæmophilic sons, X′Y, and daughters who would be apparently normal, but half of whom, XX′, would be 'carriers' of the condition but would not show the defect. Females are unaffected unless both father and mother possess the harmful gene and pass it on to the daughter, who would then have the constitution X′X′. In the case of hæmophilia this is most unlikely to occur, since a marriage between a hæmophilic man and a woman from a known family of bleeders would be very unlikely. But in the case of red-green colour blindness, which is also inherited as a sex-linked recessive, the defect occurs in about 0·4 per cent. of women in England (see p. 691).

The type of sex linkage illustrated by the inheritance of hæmophilia is that most usually met with. Genes which assort together in the same chromosome are said to be linked, and if they occur on the sex chromosome, are said to be sex-linked, although the genes in question may not be concerned with sex determination. When a gene such as that for hæmophilia is carried on the X chromosome but not on the Y chromosome, sex linkage is *total in the* X *chromosome.* If a gene is situated in the short homologous regions of the X and Y then sex linkage is only partial, as in *retinitis pigmentosa.* If the gene is in the non-pairing part of the Y chromosome, then sex linkage is *total in the* Y, but very few genes of this kind have been detected and, of course, such are transmitted down the male line only.

It has recently been claimed that it is possible to decide the sex of an individual by examination of cells from somatic tissues such as skin, buccal mucosa, polymorphonuclear leucocytes and nervous system. A little deeply staining granule, adjacent to the nucleolus, is often found in the nuclei of cells in the female but much less often in the male. The relationship between this so-called 'sex chromatin' and the sex chromosomes is, however, quite unknown, and there is a possibility of confusion of these granules with granules produced in the metabolism of the 'resting' nucleus.

It would be of great value, as well as interest, to be able to determine the sex of individuals, where this is in doubt, but it is too early yet to say how much reliance can be placed on observations on somatic nuclei.

MUTATION

On the whole the gene is an extremely stable structure, and the enzymatic process for which it is responsible is carried out with great exactitude. When a cell divides each gene first produces an exact replica of itself and one of these genes goes into each new cell. Occasionally, however, there arise changes in the arrangement or properties of the genes, and in this way mutations arise. Thus hæmophilia may appear in a member of a family which has no previous history of such an abnormality. The rate of mutation may be speeded up experimentally by treatment with certain mutagenic substances, for example mustard gas, or by exposure of the individual to radiation. At present we can only guess at the number of mutations which a given dose of X-rays will produce in human beings. It is believed that if workers in atomic establishments are not exposed to doses above 15 r per year direct radiation effects are extremely unlikely; but as far as we know even this dosage may lead to the production of undesirable mutations which may impair mental and physical abilities or even viability.

REFERENCES

BAILEY, N. T. J. (1951). Some contributions of modern genetics to medicine. *Brit. med. J.* i, 8–12.

CREW, F. A. E. (1947). *Genetics in Relation to Clinical Medicine.* Edinburgh: Oliver & Boyd.

FORD, E. B. (1946). *Genetics for Medical Students.* London: Methuen.

LAWLER, S. D. & LAWLER, L. J. (1951). *Human Blood Groups and Inheritance.* London: Heinemann.

RACE, R. R. & SANGER, R. (1950). *Blood Groups in Man.* Oxford: Blackwell.

RILEY, H. P. (1948). *Introduction to Genetics and Cytogenetics.* New York: John Wiley.

The hazards to man of nuclear and allied radiations (1956). Medical Research Council. London: H.M.S.O.

CHAPTER 55

GROWTH AND SENILITY

We have already described in various parts of this book some of the conditions necessary for normal growth. An adequate supply of food materials, including especially protein and certain vitamins, is essential but all the nutritional factors concerned in growth are not yet known. For example the difference in biological value of animal and vegetable proteins fed to animals cannot be entirely ascribed to the difference in their amino-acid content. Recent work suggests that liver and yeast both contain as yet unidentified substances (not vitamin B_{12}) which promote growth in children as well as in rats. An important factor in the internal environment of the growing organism is a properly balanced supply of hormones because the first effect of failure of almost any of the glands of internal secretion in young animals is a diminished growth rate.

Infants born at full term whose weight is under average are probably normal individuals who develop into small adults. Underweight babies are, however, usually premature, thrive less well and have a smaller chance of survival than children born at full term; it is also widely believed that infants born prematurely remain small at all ages. Recent investigations show that improvement of the diet of pregnant women decreases neonatal mortality which is due at least partly to prematurity. However, the influence of maternal nutrition on the birth weight of the child is very small for, when the mother's diet is deficient in a minor degree, the child grows *in utero* at the expense of the maternal tissues and so reaches a normal weight at birth; it is only when the mother's diet is grossly deficient that the birth weight is subnormal. On the other hand, the birth weight of the child cannot be increased by supplementing the maternal diet in any way. A factor of importance in determining the size of the offspring is the parity of the mother, for the later children in a family have a birth weight on the average 0·2 kg. greater than that of the first-born. The size of the mother has already been mentioned on p. 984 as influencing the size of the offspring so that a small mother, fortunately for herself and her baby, usually has a small child; diabetic mothers, however, no matter what their size, usually have large children.

There is no doubt that in infancy and adolescence the rate of growth has accelerated during the last forty years and this change may be attributed at least partly to improved nutrition. Whether this means that the final stature reached is now greater it is not possible to say. Experiments show that poor feeding of young animals retards development and that the full potential of growth is not attained by later improvement of the diet. In man the growth in length of the legs occurs late in adolescence and contributes considerably to the final stature and it has been said that the best fed parts of the population are relatively long limbed. Maximum height is reached in boys between 15 and 20 years and in girls between 14 and 18 years; after this increase in weight may occur. At the beginning of this century in Great Britain the weight of both men and women increased regularly with age between 20 and 60 but at the present time men on the average keep at the same weight from age 25 onwards. On the other hand women, whether single or married, with or without children, still show a gradual increase in weight from age 25 onwards.

Man reaches the zenith of his physical powers at the age of twenty years or so and sooner or later his bodily tissues, even if they do not become affected by disease, gradually begin to deteriorate. It is not at all clear why this should be so, since cells continue to grow indefinitely in tissue culture provided that the waste products of metabolism are removed. For example, Carrel has maintained cells from a chick heart growing in tissue culture for seventeen years. If these cells had been allowed to remain *in situ*, growth would have ceased when the full size of the heart had been reached as if there were some kind of inhibitory influence in the living body which limited growth. As Pearl has said, "the individual cells and tissues of the body, in and by themselves, are potentially immortal."

Although we have no explanation of the process of ageing there is no doubt that there is a limit to the length of life of mammals. We are not concerned here with the average life span, which is determined mainly by the incidence of disease or accident, rather than by physiological considerations, but with the maximum span of life which under present conditions seems to be about 110 years in man. It is not easy to find hereditary factors affecting life span since longevity may be affected by external influences such as hygiene and nutrition; the expectation of life has increased greatly in the last fifty years. In Great Britain the average life span is now 68 years as compared with 50 at the beginning of the century. It has been found that people who live to 90 or more have a greater percentage of long-lived parents than controls. The greater length of life of sons of long-lived parents as compared with sons of parents with poor records of longevity is only a matter of 2 to 4 years. There is a slight sex difference in longevity, women living on the average slightly longer than men.

McCay has shown that the life-span of rats can be greatly increased by supplying them with diets adequate in proteins, vitamins and minerals but so restricted in calorie value that the weight of the rats is maintained but not allowed to increase. Even after one thousand days on this regimen additional calories permit growth to take place at a time long after the death of all control animals born at the same time but supplied with as much food as they would eat; the kind of food supplying the extra calories has no apparent effect on the ageing process. The rats whose growth has been retarded in this way look quite young whatever their age and, provided that vitamin E is supplied, male rats retain their fertility into old age. Such animals have an increased resistance to pathogenic organisms and thus remain free from the lung diseases which are the common cause of death in rats given unrestricted diets; tumour growth is less common in the retarded animals. Experiments designed to produce the opposite effect, namely the acceleration of senescence in rats, have been uniformly unsuccessful.

If rats are fed on an adequate diet for a year and then put on a variety of experimental diets of different protein content it is found that those with normal body weight live longest irrespective of the amount of protein in the diet; overweight or underweight animals die sooner on the average.

Spells of moderate exercise such as running in rotating cages do not reduce the life-span of healthy rats. Observation of rat colonies shows that there is no correlation between the state of the ovaries or testes and the senescence of the animal as a whole. The deterioration of the gonads could hardly be responsible for the deterioration of somatic cells; if it were so senescence in eunuchs would occur unusually early.

Although every centenarian asked how he has reached his venerable age will offer advice on the kind of food to eat, it is difficult to obtain reliable information on the effect of diet on longevity in man. There is, for example, no evidence for any adverse influence of additional dietary protein on the life-span or the ageing process in man and, therefore, there is no reason on this account for restricting the protein intake of old people. Recent experiments suggest that as much as 1·4 g. of protein per kg. body weight per day may be necessary to avoid a negative nitrogen balance in old age. An adequate vitamin intake is essential for a healthy old age. An excessive caloric intake must of course be avoided by the elderly just as by the middle-aged since there is no doubt that obesity reduces the expectation of life. There may be some alteration in carbohydrate metabolism in old age, since during a glucose tolerance test (p. 288) the blood sugar often reaches a higher level and returns to the initial value more slowly than in young adults.

Senility is not a state readily capable of definition nor indeed is there a very clear pathological basis for it, since the examination of the bodies of persons who have died of old age does not reveal gross abnormality in any vital organ. The main characteristic of senility is atrophy of organs as shown by a decrease of weight and this weight loss may be used as an index of ageing. Atrophy of one organ may affect other organs which depend on it and so accelerate the process of senescence. Damage of an organ by infection or by any pathological process may similarly speed up the ageing process. It is often stated that in old age there is a dehydration of all tissues but recent investigations show exactly the reverse, namely that there is on the average an increase in the water content of the tissues of old people. This must be due to an increase of extracellular fluid, since the content of sodium, a characteristically extracellular constituent, is raised and the content of potassium, an index of intracellular water, is lowered in aged tissues.

The interpretation of biochemical and physiological measurements in old people must be made with care, since the range of individual variation is much greater than in younger persons and this is especially true of the basal metabolic rate. On the average the B.M.R. of old people is slightly less than that of young adults (Table 11, 16) and this reduction is presumably a reflexion of the diminution of active cellular material in the body or, to put the matter another way, it may be brought about by the increase of relatively inert material such as collagen, fat and fibrin in the tissues of the aged. It has been found that there is a decline in the rate of uptake of oxygen of tissues of old animals as compared with that of tissues from young animals. This may be due to a fall in the activity of the intracellular enzymes such as the cytochrome oxidase systems.

Changes in individual organs with age have already been mentioned in previous chapters. We judge the age of a person mainly by the appearance of his skin which in old age shows wrinkling, thinning and loss of elasticity. The electron microscope shows that the collagen network in the skin of exposed areas in old people is defective, the fibres being short, bent and frayed. The collagen of unexposed sites is normal. Although gastric acidity declines with old age and achlorhydria is common (p. 248), absorption is good and an old person in good health utilizes food as well as a younger person. The changes in the elasticity of the arteries are illustrated on p. 481. The calcium content of arteries and cartilage increases in old age but there is no alteration in the blood calcium level. The cholesterol content of arterial walls, of the cornea and the lens increases often without any obvious change in blood cholesterol. The percentage of ash in bones tends to rise in old age and the marrow cavity may become enlarged; X-ray photographs of elderly persons show that there is, especially in women, an alteration in the architecture which has been interpreted as being due to osteoporosis but objective measurements of the density of the shadows often fail to distinguish between, for example, the vertebral columns of young and aged subjects. The healing of fractures in the aged may be greatly slowed down. Old people are often in negative calcium balance. The presbyopic changes are described on p. 676. Renal function becomes progressively less efficient in old age for the blood urea tends to be higher than normal and the excretion of injected dyestuffs to occur more slowly.

Tests of muscle strength show that the maximum is reached somewhere about 20 years of age and is retained to at least 40 years; thereafter there is a slight decline, which becomes marked only after 70 years, when the maximum work output may be only 50 per cent. of that of the young adult. These findings may be related to the histological changes in the muscles of old people which are described on p. 771. Increasing difficulty in carrying out physical exercise is a well-known characteristic of old age but there may be no change in the apparent efficiency (p. 164) with which muscular work of moderate severity is performed.

Thus it can be said that the changes of old age are not very sudden or very striking. It would appear that the organs of the body wear out slowly and that death is determined partly by a progressive atrophy of all the body tissues and partly by the gradual failure of some vital organ. The idea that senescence is due to 'wear and tear' may be true but it is difficult to prove and what 'wear and tear' means depends on the expounder of the notion and his experience. Neurones cannot apparently produce new nerve cells yet they may survive actively in man for over a hundred years, but in other mammals for a shorter period. There is no way of finding the life span of cells in the tissues of the living body but cell replacement, as shown by mitotic figures, goes on in many tissues for example, skin, liver, adrenal cortex, and anterior pituitary; this activity declines in senescence. Senescence might be brought about by the exhaustion of certain essential cell constituents, such as enzymes and there is some evidence in favour of this. The limited survival of red cells may be due to the damaging effect of the products of one enzyme system on others operating in the cell. There is evidence, such as the decline in the output of 17-ketosteroids (17-oxosteroids) in old age, that some at least of the endocrines decline in activity. The cause of this is not known and treatment by replacement, as the administration of œstrogens after the menopause, has only a limited rejuvenating effect. It is perhaps too much to expect a single cause of senescence and a solution to the problem of the cause of senescence and of prolongation of life is not likely to be forthcoming for a long time. In the meantime many urgent practical problems, psychological, social, and medical are being thrust upon us by an increasingly aged population.

REFERENCES

Andrew, W. (1952). *Cellular Changes with Age.* Springfield, Illinois: Thomas.

Comfort, A. (1956). *The Biology of Senescence.* London: Routledge & Kegan Paul.

Cowdry, E. V. (1952). *Problems of Ageing.* 3rd ed., ed. A. I. Lansing. Baltimore: Williams & Wilkins.

Ellis, R. W. B. (1951). Phases of postnatal growth. *Brit. J. Nutr.* **5,** 151–157.

Engle, E. T. & Pincus, G. (Eds.) (1956). *Hormones and the Aging Process.* London: Baillière, Tindall & Cox.

Heller, A. L. & Shipley, R. A. (1951). Endocrine studies in ageing. *J. clin. Endocr.* **11,** 945–962.

Korenchevsky, V. (1949). The problem of ageing. *Brit. med. J.* i, 66–68.

Lansing, A. I. (1951). Some physiological aspects of ageing. *Physiol. Rev.* **31,** 274–284.

Leitch, I. (1951). Growth and health. *Brit. J. Nutr.* **5,** 142–151.

Lowe, C. R. & Gibson, J. R. (1955). Changes in body weight associated with age and marital status. *Brit. med. J.* ii, 1006–1008.

McKeown, T. & Record, R. G. (1952). Observations on fœtal growth in multiple pregnancy in man. *J. Endocr.* **8,** 386–401.

Medawar, P. B. (1945). Old age and natural death. *Modern Quart.* **1,** 30.

Shock, N. W. (1948). Metabolism in old age. *Bull. N.Y. Acad. Med.* **24,** 166–178.

Simpson, S. L. (1950). Physiology of growth. *Postgrad. med. J.* **26,** 417–424.

Tanner, J. M. (1955). *Growth at Adolescence.* Oxford: Blackwell.

Thomson, A. M. (1951). Human fœtal growth. *Brit. J. Nutr.* **5,** 158–166.

Walker, K. (1952). *Commentary on Age.* London: Case.

Weir, J. B. de V. (1952). The assessment of the growth of school children with special reference to secular changes. *Brit. J. Nutr.* **6,** 19–33.

UNITS AND MEASURES (WITH CONVERSION FACTORS) COMMONLY USED IN PHYSIOLOGY AND BIOCHEMISTRY

Prefixes applied to symbols for various units: these are never used alone

M = Mega = 10^6
k = kilo = 10^3
m = milli = 10^{-3}
μ = micro = 10^{-6}
n = nano = 10^{-9}
p = pico = 10^{-12}

Mass

gram	g.
kilogram	kg.
milligram	mg.
microgram	μg.

1 kg. = 1000 g.
1 g. = 1000 mg. = 1,000,000 μg.
1 kg. = 2·20 lb.
1 g. = 15·4 grains
1 g. = 0·035 oz
1 lb. = 453·6 g.
1 oz = 28·4 g.
1 grain = 65 mg.

Length

metre	m	micron	μ
centimetre	cm	millimicron	mμ
millimetre	mm	Angstrom unit	Å

1 m = 100 cm = 1000 mm = 10^6 μ = 10^9 mμ = 10^{10} Å
1 μ = 10^{-6} m = 1000 mμ = 10,000 Å
1 mμ = 10^{-9} m = 10 Å
1 cm = 0·394 in.
1 kilometre = 0·62 mile.
1 in. = 2·54 cm
1 mile = 1·61 kilometres

Time

hour	hr
minute	min
second	sec
millisecond	msec (occasionally written σ)

1 hr = 60 min = 3600 sec
1 sec = 1000 msec = 1,000,000 μsec

Volume

litre	l.
millilitre	ml.
microlitre	μl.
cubic centimetre	c.c.

The volume of 1 kg. of water at the temperature of maximum density (4° C) is 1 l. The millilitre (ml.) is one thousandth part of a litre. The cubic centimetre (c.c.) is based on the standard of length, that is the metre, and is very nearly but not the same as the ml. (see below).

1 l. = 1000 ml.
1 l. = 35·196 fluid ounces (oz)
1 l. = 1·76 pint
1 pint = 568 ml.
1 ml. = $\frac{1}{1000}$ l. = 1·000027 c.c.
1 ml. = 1000 μl.
1 oz = 28·4 ml.
1 gallon = 4·55 l.

Heat

The calorie (cal), also known as the small calorie or gram calorie, is the amount of heat required to raise the temperature of 1 g. of water from 3·5° C to 4·5° C.

The large calorie (kcal), also known as the kilogram calorie or kilocalorie, is the amount of heat required to raise the temperature of 1 kg. of water from 3·5° C to 4·5° C.

1 kcal = 1000 cal = 4186 joules.

Conversion from Fahrenheit to Centigrade scale of temperature or vice versa may be made by using the formula—

$$9/5\ C = F - 32$$

where F is the temperature in degrees Fahrenheit
and C is the temperature in degrees Centigrade.

Electrical Units

ampere, A
milliampere, mA
watt, W
farad, F
microfarad, μF
coulomb, C
volt, V
millivolt, mV
ohm, Ω
megohm, MΩ
joule, J

Miscellaneous

cycles per second, c/s
normal temperature and pressure, s.t.p.
acceleration due to gravity $= g$

Unit of force is the dyne or dyn. One dyne is the force which would produce an acceleration of 1 cm/sec/sec if acting on 1 g.

Unit of work is the erg. One erg is the work done when a force of one dyne acts through one cm in the direction of the force.

1 joule (J) $= 10^7$ ergs.

Power is the rate of doing work.

1 watt $=$ 1 joule per sec.

1 horse power (h.p.) $=$ 33,000 ft. Lb./min $=$ 746 W.

STANDARD WEIGHTS, ACCORDING TO SEX, AGE AND HEIGHT

1. Men

Age Period	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–	Age Period
Height	lb.	lb.	lb.	lb.	lb.	lb.	lb.	lb.	lb.	Height
ft. in.										cm
4 11	111	117	122	125	127	130	132	133	134	150
5 0	113	119	124	127	129	132	134	135	136	152
5 1	115	121	126	129	131	134	136	137	138	155
5 2	118	124	128	131	133	136	138	139	140	157
5 3	121	127	131	134	136	139	141	142	143	160
5 4	124	131	134	137	140	142	144	145	146	163
5 5	128	135	138	141	144	146	148	149	150	165
5 6	132	139	142	145	148	150	152	153	154	168
5 7	136	142	146	149	152	154	156	157	158	170
5 8	140	146	150	154	157	159	161	162	163	173
5 9	144	150	154	158	162	164	166	167	168	175
5 10	148	154	158	163	167	169	171	172	173	178
5 11	153	158	163	168	172	175	177	178	179	180
6 0	158	163	169	174	178	181	183	184	185	183
6 1	163	168	175	180	184	187	190	191	192	185
6 2	168	173	181	186	191	194	197	198	199	188
6 3	173	178	187	192	197	201	204	205	206	190

2. Women

Age Period	15–19	20–24	25–29	30–34	35–39	40–44	45–49	50–54	55–	Age Period
Height	lb.	lb.	lb.	lb.	lb.	lb.	lb.	lb.	lb.	Height
ft. in.										cm
4 11	110	113	116	119	122	126	129	131	132	150
5 0	112	115	118	121	124	128	131	133	134	152
5 1	114	117	120	123	126	130	133	135	136	155
5 2	117	120	122	125	129	133	136	138	140	157
5 3	120	123	125	128	132	136	139	141	143	160
5 4	123	126	129	132	136	139	142	144	146	163
5 5	126	129	132	136	140	143	146	148	150	165
5 6	130	133	136	140	144	147	151	152	153	168
5 7	134	137	140	144	148	151	155	157	158	170
5 8	138	141	144	148	152	155	159	162	163	173
5 9	141	145	148	152	156	159	163	166	167	175
5 10	145	149	152	155	159	162	166	170	173	178
5 11	150	153	155	158	162	166	170	174	177	180
6 0	155	157	159	162	165	169	173	177	182	183

HEIGHT: without shoes. WEIGHT: indoor clothes without shoes. For naked weight subtract 10 lb. for men and 6 lb. for women.

The accompanying table of standard weights for sex, age and height has been borrowed from Davidson, L. S. P., *Principles and Practice of Medicine* (1952), and this author's comments on it may be summarized by saying that no really satisfactory tables exist but those shown have been used with apparent satisfaction for half a century by insurance companies throughout the English-speaking world.

The values in the table are based on figures obtained at the end of last century and show an increase of weight with age. While a substantial increase in weight amounting to obesity is associated with an increase in morbidity and mortality it would be difficult to claim that the small increase shown here could impair health. Recent investigations show that the tendency to gain weight with age is on the average smaller than it was fifty years ago.

INDEX

- Abdomen, X-ray photograph after barium meal, (fig.) *facing page* 232
- Abdominal pressure in respiration, 544
- Abortion, pregnanediol excretion and, 990
- Absorptiometers, 406
- Absorption from lungs, 527
  - in colon, 279
    - large intestine, 279
    - mouth, 268
    - small intestine, 268
    - stomach, 268
  - intestinal, 268, 269
  - of carbohydrates, 269
    - fat, 270
    - iron, 392
    - minerals, 269
    - protein and amino-acids, 269
    - sterols, 275
    - vitamin K, 275
  - selective, 268
  - spectra, 83, 408
    - of blood pigments, 408
    - of hæmoglobin derivatives, (fig.) *facing page* 408
- Acceleration, effect on circulation, 492
- Accessory food factors, 109
- Acclimitization, 558
- Accommodation, 674, (fig.) *facing page* 681
- Accommodation, atropine, 675
  - eserine, 675
  - in nerve, 755
  - pathway, 675
  - range of, 676
- Acetic acid and detoxication, 419
- Acetone, 332
  - bodies, 332
  - in estimation of circulation time, 504
- Acetoacetic acid, 283, 330
- Acetylation, 420
- Acetylcholine, 221, 233, 242, 267, 456, 459, 748, 788, 813, 945
  - and gastric juice, 242
    - intestinal villi, 263
    - liberation of histamine, 242
    - salivary secretion, 220
    - stomach, 242
  - cholinergic nerves and, 812
  - contraction of intestine and, 260
  - detection of, 813
  - formula, 813
  - iris and, 677
  - muscle contraction and, 788
  - on intestine, 260
  - transmitter of effects of nerve impulses across synapse, 748
  - vagus nerve and, 459
  - vasodilatation in blood vessels of lungs, 528
  - vasodilator substance, 516
- Acetyl coenzyme A, 316, 330, 748
  - phosphate, 61
- Achlorhydria, 248
  - and absorption of iron, 392
  - incidence of, 248
- Achromotrichia, 130
- Acid hæmatin, 406, 408
  - absorption spectra, *facing page* 408
  - phosphatase, 104
- Acid-base and the kidney, 620
- Acidæmia, 641
- Acidophils, 998
- Acidification of urine, 621
- Acidosis, 641, 643
  - carbon dioxide, 642
  - excessive loss of base, 644
  - from excess of acids other than carbon dioxide, 642
  - impaired elimination of acid, 644
  - in diabetes mellitus, 310, 332, 644
- Acinus, 219
- Aconitase, 317
- Acoustic spectrum, 709
- Acrodynia, 130
- Acromegaly, 1002, 1007, (fig.) *facing page* 1008
  - metabolic rate and, 1008
- ACTH (adrenocorticotrophic hormone), 1004
  - assay of, 1004
  - milk secretion, 994
- Actin, 773
- Action potential, 736, 737, 743
  - diphasic and monophasic, 742
- Activators, enzyme, 75
- Actomyosin, 773, 963
- Acuity of hearing, 724
- Adaptation, 3
  - and taste, 668
  - definition of, 3
  - in olfaction, 663
    - pain sense, 658
    - smell sense, 663
    - taste sense, 668
    - temperature sense, 652
  - of end organs, 768
  - of tactile sense, 655
  - to hot and cold, 653
- Addison's disease, 201, 619, 939, 953
  - DCA in, 940
- Adenase, 355
- Adenine, 57, 65
- Adenocorticotrophic hormone (ACTH), 1003, 1004
- Adenohypophysial hormones (table), 1003
- Adenohypophysis, 998, (fig.) *facing page* 1001, 1002
  - and sexual activity, 970
  - basophils and acidophils, control of, 970
  - blood supply, 1000
  - clinical disorders of, 1007
  - diabetogenic hormone, 1005
  - divisions of, 998
  - functions of, 1001
  - growth hormones, 1005
  - hormonal transmitter, 1001
  - hormones of, (table) 1003
  - hypofunction of, 1004
  - lactogenic hormone, 993, 1006
  - melanophore-expanding hormone, 1007

Adenohypophysis, œstrous cycle, 970
  prolactin, 993
  structure, 999
Adenosine, 59
Adenosinediphosphate (ADP), 60, 776
Adenosinetriphosphate (ATP), 773
  and myosin, 773
  in muscle glycolysis, 777
  in muscular contraction, 776
Adenyl pyrophosphate, 60
Adenylic acid, 60
  a vasodilator substance, 516
Adiadochokinesis, 875
Adolescents, height, weight and surface area, 167
Adrenal cortex, 28, 305, 936
  androgenic material excreted in urine and, 953
  androgens in, 938
  and water balance, 619
  anterior pituitary and ACTH, 1004
  ascorbic acid in, 936
  cortical hormones, 938
  cortical tumour and, 941
  extirpation of both suprarenal glands, 936
  extracts, action of, 938
  fat metabolism and, 272
  general adaptation-syndrome, 939
  glands, 935
  hormones of, 937
    and carbohydrate metabolism, 305
  hyperfunction, 941
  in carbohydrate metabolism, 305
  medulla, 304, 941
    hyperfunction, 944
  nerve supply, 941
  œstrogens, 938
  stress, 943
  tumours and paroxysmal hypertension, 944
  venous effluent, 942
Adrenalectomy, 272, 937
Adrenaline, 221, 234, 304, 306, 397, 449, 516, 532, 546, 678, 810, 941
  and blood lactate, 299
    bronchial muscle, 546
    cerebral arteries, (fig.) 534
    glycogenolysis, 305
    heart output, 463
    metabolic rate, 201
    noradrenaline, effects of, 811
    phosphorylation, 294
    salivary secretion, 221
  apnœa, 564
  asthma and, 546
  blood-sugar level and, 306
  capillary contraction and, 518
  cardiac output and, (fig.) 463
  denervation and, 811
  destruction of, 942
  direct action on blood vessels, 517
  discharge from suprarenal medulla, 942
  discrepancies between effects of injection of and sympathetic stimulation, 810
  effect on blood pressure, (fig.) *facing page* 521
  effects of, (table) 811, (fig.) *facing page* 941
  emergency function, 943
  estimation of, 942
  gastric contractions and, 234
    juice and, 242
Adrenaline, gastric motility and, 234
  glycogenolysis in liver and muscle, 304
  glycosuria, 306
  in asphyxia, 515
    blood, 944
  infusion, 463
  intravenous injection in animals and arterial blood pressure, 516
  iris and, 678
  liberated at low blood-sugar level, 302
  metabolic rate and, 201
  noradrenaline, action of, compared with adrenaline, (table) 811
  on blood pressure, 516
    heart, 462
    intestine, 260
    spleen, 378
  pressure in the carotid sinus and, 516
  reflex effects on alteration of pressure within carotid sinus, 464
  respiration and, 564
  stimulation of splanchnic nerves, 516
  sympathetic nervous system and, 462, 805, 810
  synthesis of, 942
  urinary output, 943
  vasoconstriction in blood vessels of lungs, 529
  X-ray examination of spleen and, 532
Adrenergic nerves, 753, 810, diagram 814
Adrenocortical virilism, (fig.) *facing page* 940
Adrenocorticotrophic hormone (ACTH), 301, 939
  assay, 1004
  potentiates growth hormone, 301
  structure, 1005
  adrenogenital syndrome, 941
Aerobic dehydrogenases, 81
  delayed heat, 791
  oxidases, 81
  recovery phase in muscle, 777
Aerophagy, 227
Aesthesiometer, 656
Ætioporphyrin, 412
Afferent fibres, 817
  impulses, 765
    from viscera, 805
  nerves, definition, 5
Afibrinogenæmia, 368
After-discharge, 833
After-image, 691
After-potential, 745
Age and basal oxygen requirement, (table) 544
  blood pressure, 487
  pulse rate, (fig.) 467
  respiratory rate, (table) 544
Ageing, process of, 1022
Agglutination, (table) 380
  differential, in estimation of life of red cells, 391
Agglutinins, 380
Agglutinogens, 380
Aglomerular kidneys, 600
Air, alveolar, 551
  atmosphere, 552
  breathed, volume of, 547
  comparison of expired, alveolar and inspired, (table) 552
  composition of the respired, 551
  conduction, 727

Air, dead space, 549
  expired, 551
  functional residual, 548
  hunger, 642
  passages, 537
  residual, 548
  respired, composition of, 547
  tidal, 548
  vital capacity, 548
  volume breathed, 548
Aircraft, 492
Alanine, 40
Albumin, 369
  in osmotic equilibrium, 372
  plasma protein, 372
  replacing blood plasma in blood transfusion, 374
Albumins, 53, 371
Alcohol, 182, 314
  calorific value of, 182
  thirst and, 217
Alcoholic fermentation, 314
  the Pasteur effect, 316
Aldoses, 8
Aldosterone, 28, 619, 938, 939
Aleurone layer, 178
Alkalæmia, 557
Alkali reserve, 641
  tetany, 931
Alkaline phosphatase, 104, 257, 397
  in bile, 257
  tide, 240
Alkaloidal reagents, 54
Alkalosis, 641, 643
  intake of alkali, 644
Alkaptonuria, 139, 284, 349
Alipotropic action, 327
Allantoin, 356
Allelic pair, 1016
'All-or-nothing Law,' 428, 739, 781
  in muscle, (fig.) *facing page* 784
Alloxan, 58, 307
  glycosuria, 307
Alloxazine adenine dinucleotide, 62, 87
  mononucleotide, 86
Alternating currents, 757
Altitude, high and respiration, 557
Alveolar air, 551, 556, 583
  at different pressures, (table) 556
  apparatus for collecting, (fig.) 552
  membrane, 539, 582
Alveolar capillary, corpuscle in, 517
Alveolar membrane, diffusing capacity, 584
    passage of the gases through, 582
Alveoli, 538
Alveolus, salivary, 219
Amine oxidase, 82, 811
Amino-acetic acid, 40
Amino-acid, 40
  decarboxylase, 129
  oxidase, 341
Amino-acids, absorption of, 269
  aliphatic, 41
  and nitrogen balance, 184
    urea formation, 342
  antiketogenic or glucogenic, 345
  aromatic and heterocyclic, 42
  breakdown, 339
  classified, 41
  conjugation, 347
Amino-acids, diamino-monocarboxylic, 42
  effect on weight of rat, (fig.) 185
  essential, 184
    for adult man, 185
    (ten) for growth of young rats, 185
  fate after absorption, (table) 339
  in blood, 335
  in protein, 47
  in tissues, 337
  indispensable, 184
  intravenous infusion, 339
  ketogenic and antiketogenic, 345
  list of, 41–43
  mono-amino-dicarboxylic, 41
  nitrogen isotope ^{15}N, 338
    balance, 185
  oxidase, 82, 87, 341
  plasma protein formed and, 373
  protein metabolism, 335–359
  reactions, 347
  storage, 335
  sulphur containing, 354
  synthesis of by body, 184
  transamination, 341
  transmethylation, 347
  urea from, 342
Amino-aciduria, 357
p-aminobenzene sulphonamide (sulphanilamide), 134
p-aminobenzoic acid, 134
  an essential metabolite, 134
  folic acid and, 134
  relationship to sulphonamides, 134
p-aminohippurate, 623
p-amino hippuric acid, 606
Aminoethyl alcohol, 23
Amino-nitrogen absorption curves, (fig.) 335
Aminopeptidases, 265
  in intestinal juice, 265
    pancreatic juice, 252
Amino-propionic acid, 40
Amino-sugar, 12
Ammonia formation, 345
  in urine, 346
AMP, adenosine monophosphate, 61
Amœboid movements, 397
Amphetamine and hunger, 215
Ampholytes, 43
Amphoteric electrolytes, 43
Amniotic fluid, 976
Amplitude of accommodation, (fig.) 676
Ampulla of the bile duct, 251
  semicircular duct, 856
Ampullary crests, 856
Amylase, 69, 222, 265, 294
  in intestinal juice, 264
  pancreatic, 253
  salivary, 222
Amylopectin, 16
Amylose, 16
Anabolic processes, 2
Anæmia, 386, 394, 586
  Addisonian, *facing page* 388
  anæmic anoxia, 586
  hæmoglobin colour index (C.I.) in, 406
  hæmopoiesis and, 395
  nutritional, 99
  pernicious, *facing page* 388
    liver treatment in, 394
Anærobic delayed heat, 790

Anæsthesia, 754, 830
Anæsthetic drugs, 749
Anal sphincters, 278
Analyser, 915
Analysis of gastric contents, 246
Anaphylactic reaction, 56
Anaphylaxis, 56
Anastomosis, arteriovenous, 206, 231, 263, 456, 519, *facing page* 520, (fig.) *facing page* 521
Androgens, 27, 952
  amount of androgenic material excreted in urine, (table) 953
  and comb growth, 952
  and pituitary inhibition, 970
  and protein anabolism, 954
  biological test for, 951, 954
  effect on protein anabolism, 954
  excretion, 952
  in adrenal cortex, 938
Androstane, 27
Androsterone, (formula), 951
Anelectrotonus, 754
Aneurine, 121
Aneurism, 501
Angina pectoris, 808
Angio-cardiography, 424
Angiostomy, 284
Angular acceleration, 858
Angular gyrus, 904
  stomatitis, 126
Anion exchangers, 38
Annulospiral endings, 772
Anœstrus, 961
Anovular menstruation, 959
Anoxæmia, 585
Anoxia, 585, 942
  anoxic, 585
  colour of skin in, 587
  flying at high altitudes and, 589
  histotoxic, 587
  stagnant, 587
Anoxic anoxia, 585
Anterior lobe of pituitary, 300, 998
  poliomyelitis, 828
  root section, 828
Antibiotics, 45
Antibodies, 55, 193, 378
  Rh, 382
Anticholinesterase, 74, 221, 233, 678, 789, 813
Anticoagulants, 362
Antidiuretic effect, 1009
  hormone (ADH), 613, 616, 618
  with cardiac failure, and in liver disease, 1010
Antidromic impulse, 739
  vasodilatation, 511
Anti-enzymes, 75
Antigens, 55, 378
Antihæmophilic globulin, 365, 366
Anti-histamine substances, 658
Antiketogenesis, 332
Antiketogenic amino-acids, 333, 345
  and glycogen, 297
Antilipotropic action, 327
'Anti-oxidants,' 21, 116
Antiperistaltic movements, 262
Antiplasmin, 367
Antipernicious-anæmia principle, 394
Antipyrine, 630
Antithyroid substance, 825
Antitoxins, 378
Antivitamins, 130, 134, 135
Antrum of stomach, 243
  and gastric secretion, 244
Aorta, volume-elasticity curves, 481
Aortic chemoreceptors, 513
Aortic pressure, (fig.) 434
Apatite, 104
Apex beat, 438
Aphasia, 902, 904
Apneusis, 554
Apnœa, 557
  voluntary, 557
Apocrine glands, 652
Apoenzyme, 76, 77
Apoferritin, 392
Apomorphine, 249
Appendicitis, 808
Appendix, vermiform, 276
Apperception, 648
Appetite, 202, 215
  juice, 242
Aqueous humour, 671
Arabinose, absorption of, in intestine, 269
Arachnoid mater, 644
Arbor vitæ, 865
Archipallium, 887
Architectonic fields, 888
Arciform arteries, 591
Area, surface, 163
Arginase, 77, 342
Arginine, 342, 352
  and spermatozoa, 186
  essential amino-acid, 185
Ariboflavinosis, 126
Aristotle's experiment, 656
Arm-lift-back-pressure method of artificial respiration, (fig.) 567
Arrhythmia, respiratory, 473, 495
Arterenol, 941
Arterial blood gases, 572, 583
Arterial pressure, (fig.) *facing page* 457, 483, 491
  age variations, 487
  and heart output, (fig.) 444
  and respiration, (fig.) *facing page* 529
  auscultatory method of measuring, 484
  brachial and femoral arteries, 487
  in animals, 483
  in heart-lung preparation, 442
  mean, 484
  peripheral resistance and mean, 499
  pulse, 493
  systolic blood pressure (S.B.P.) normal, 487
Arteries in old age, 1024
Arterioles, glomerular, 596
Arteriovenous anastomosis, 231, 263, 519, (fig.) *facing page* 520, (fig.) *facing page* 521
  heart, 456
  of skin, 206
Artery, (fig.) *facing page* 480
  and vein compared, (fig.) *facing page* 480
  brachial pressure, (fig.) *facing page* 528
  elasticity, (fig.) 481
  pulmonary, pressure change due to intrathoracic pressure, 529
  structure of, 480, (fig.) *following page* 480
  volume-elasticity curves, 481

Artificial pyrexia, 610
Artificial respiration, 566
  Holger-Nielsen method, 567
  rocking method, 569
Ascending degeneration in cord, (fig.) 825
Ascending paths in spinal cord, 825
Aschheim-Zondek test, 978, 979
Ascorbic acid, 136, 672
  and anæmia, 395
    bones, 139
    oxidase, 81, 82, 138
    teeth, 140
    wounds, 140
  daily intake of, 142
  deficiency and wound healing, (fig.) *facing page* 140
  destroyed by cooking, 138, 141
  estimation of, 137
  hydrogen carrier, 138
  in blood and tissues, 138
  in foods, (table) 141
  in potatoes, 141
  in suprarenal, 936, 942
  in tissues, 138
  in white blood cells, 139
  loss from cooked cabbage, (table) 142
  of adrenal cortex, 1004
  stability of, 138
  vitamin C, 136–142
Aspartic acid, 41
Asphyxia, 514, 555
  and pain sense, 648
  on blood pressure, (fig.) *following page* 512
  respiratory movements in, 555
Association fibres, 888
Asthenia, 870
Asthma, 915
Astigmatism, 680
Astronomer's trick, 697
Asymmetric carbon atoms, 9, 40
Ataxia, cerebellar, 870
Ateleiosis, 1008
Atherosclerosis, 333
Athetosis, 886
Athletes, 459
Atmospheric air, 552
ATP, 60, 61, 313, 314, 318, 753
ATP and muscle contraction, 775
Atrial extrasystole, 474
  fibrillation, 473, 475, 477, 478
  flutter, 473, 475, 477, 478
  pressure, 433, (fig.) *facing page* 488
  pressure and heart output, 451
    man, 433, 451
    right and left, (fig.) *facing page* 457
  reflex, right, 466
Atrioventricular bundle, 453
  node, 424, 425
  valves, 425
Atropine, 205, 216, 220, 221, 233, 253, 459, 678, 788
  and gastric motility, 233
  and heart rate, 459
    iris, 678
    pancreas, 253
    salivary secretion, 220
    sweat, 205
    thirst, 216
  preventing effect of vagal impulses on the heart 459
Atwater-Benedict respiration calorimeter, 158
Atwater values, 152
Audibility limits, 725
Audibility, threshold, 724
Audiogram, 727
  chart, 726
Audiometer, gramophone, 728
  pure tone, 724
Auditory area, 902
  cortex, 725, 901
  nerve, 719
  action potential of, 718
  sensation area, 724
Auerbach's plexus, 230
Auricle, 711
Ausculatory method of determining arterial pressure, 484
Autolysis, 78
Autonomic activities, cortical representation of, 896
  fibres to eye, 901
  nervous system, 804
  outflow, (fig.) *facing page* 804
Axis cylinder, 733
Axon, 733
  reflex, 512
Axonotmesis, 763

Babinski response, 830, 894
Baby, iron content of, 97
  growth curve of, 994
Bacteria of intestine, 280
Bacterial synthesis of vitamins, 280
Bag of membranes, 976
Bainbridge reflex, 466
BAL, British anti-lewisite, 74
Balance technique, 96
Baldness, 951, 954
Ballistocardiogram, 438
Balloon method of recording gastric contractions, 232
Band of Baillarger, 888
Barium enema, 276 (fig.) 279
  meal, (fig.) *facing page* 232, (fig.) *facing page* 233
Baroceptors, 463, 513
Basal ganglia, 884
  metabolic rate, 165, 201
    and environmental temperature, 201
    at different ages, (table) 168
    methods of determination, 157
    specific dynamic action and, 169
    thyroid gland activity and, 924
  nuclei, 884
  oxygen requirement, 544
Basilar membrane, 714, 720
  displacement of, 716
  travelling wave, 716
Basophil cells of the adenohypophysis, 1008
  leucocytes, 945
Basophilic adenoma, 1004
Basophilism, 1008
Basophils, 397, 998
Beer, 182
Beer's Law, 406
Beeswax, 21
Beet sugar, 14
Belching, 227 708
Bends 560

Benedict-Roth apparatus, 159
Benedict's solution, 11
Benzidine, 83
  test for blood, 407
Benzoic acid, 347, 419
  and liver function, 419
Beri-beri, 109, 121, 123
Bernouille's theorem, 501
Betaine (formula), 327
Betz cells, 890
Bicycle ergometer, (fig.) 164
  pedalling, efficiency of, (fig.) 784
Bier's spots, 516
Bile, 255
  acids, 25
  alkaline phosphatase in, 257
  bile salts, 117, 220, 225, 255, 270
    hydrotropic power of, 256
  composition of human, 255
  duct, 251
  pigments, 256, 645
  pressure, 258
  white, 257
Bilirubin, 415
  concentration in blood, 415
Biliverdin, 415
Binaural masking, 729
Binocular vision, 700
Biochemical lesion, 113
Biochemistry, definition, 1
Biological oxidations and reductions, 80
  reactions of proteins, 55
  value, 189
Biotin, 131
Birth, change in circulation at, 986
  weight, 984
  factors in, 1021
Bitter taste, 667
Biuret test, 55
Black Hole of Calcutta, 213
Black-tongue, 127
  and nicotinic acid, 127
Bladder, innervation, (fig.) 625
  in paraplegia, 627
  pressure, (fig.) 626
  urinary, 830
Blastocyst, 976
Bleeder, 366
Bleeding time, 368
Blind spot, 685
Blindness, 898
  and removal of occipital cortex, 844
Blinking, 670, 671
Block, neuromuscular, 787
Blocking drugs, 748
Blood, 360–384
Blood, amino-acids in, 293
  buffering mechanism, 641
Blood-brain barrier, 645
Blood calcium, 875
Blood, carbon dioxide dissociation curves, (fig.) 579
Blood cells, 385
Blood cells, origin of, 388, 389
  altitude and number, 386
Blood cells, white, 396
  enumeration of, 399
  life of, 397
Blood chemistry, 360
  cholesterol content, 333, 360
Blood-clotting mechanism, 361
  fibrinolysis, 367
  hæmophilia, 366
  platelets and, 366
  scheme of, 365
Blood coagulation, 361
  anticoagulants, 362
  clotting time, 362
  vitamin K deficiency and, 148
Blood compatibility, 381
Blood, composition, 360
Blood constituents, concentration, 366
Blood defibrinated, 362
Blood, distribution of, in organs, 378
  changes from arterial to venous, (fig.) 583
  donors, 523
Blood, estimation of oxygen and carbon dioxide in, 572
Blood flow, 501
  in liver, 529
  loss and syncope, 523
  measurement of, by Ludwig's stromuhr, 502,
    in limb, by plethysmograph, 503
  observation in blood vessels, 505
  through brain, 820
  through forearm, (fig.) *facing page* 520
Blood formation, 392
  copper, 395
  protein, 395
Blood, formed elements of the, 385–403
Blood gas apparatus, 572
Blood gases, 572
Blood groups, 379
    forensic significance, 1018
    identity and, 1019
    inheritance of, 1017
    Rh factor, 382
  hydrogen-ion concentration, 641
  in large and minute vessels, (fig.) 379
  in lungs, measurement of, 528
  lactic acid, 799
  oxygen dissociation curves, 575
  physiological changes in infancy and childhood, 402
  pigments, 404–416
    absorption spectra, 408
Blood in infancy and childhood, 402
Blood, iodine, 923, 924
Blood, phosphorus in, 103
Blood plasma, carbon dioxide and, 578
  cerebrospinal fluid and, chemical composition compared, (table) 646
  lipid concentration, 324
  proteins of, 369
  ratio to corpuscles, 377, 981
  urine and composition of, 593, 603
  volume, measurement of, 376
Blood platelets, 368, 400, 946
Blood pressure and age, 487
  apparatus for recording, (fig.) 483
  asphyxia, (fig.) *following page* 512
  basal, 486
  cerebral, 533
  diastolic (D.B.P.), 485
  factors governing, (fig.) 492
  in animals, 483
  kidneys and, 610
  mean, 491
  relation between pressure and velocity, 493

Blood pressure, relation of the kidneys to, 610
  splanchnic nerves and, 510
  systolic (S.B.P.), 485
  vagus on, (fig.) 460
  vasomotor reflexes and, 512
Blood proteins, functions of, (table) 374
  regeneration, 373
  specific gravity of, 375
Blood, red cells, erythropoiesis in bone marrow, 389
  hæmolysis, 387
  high altitudes, increase at, 386
  life of, and methods of estimation, 390, 397
  mean corpuscular volume, 387
  number, 386
  origin of, (table and fig.) 388
  osmotic pressure and, 387
  sedimentation rate and, 389
Blood serum, 360, 362
  stagnation of in shock, 524
Blood-sugar, 288, 297
  after glucose, 289
    several doses of glucose, (fig) 291
  and adrenaline, 302
    hunger, 215, 297
    insulin, (fig.) 302, 303
  curves, (fig.) 289
    in diabetes, (fig.) 291
  effect of liver on, (fig.) 296
    meals on, 289
  gastric juice and, 248
  hormonal control of, 300
  in diabetes mellitus, effect of insulin, (fig.) 298
  origin of, 297
Blood, tests for, 407
  transfusion, 379
    donors, fainting in, 523
  velocity, 502
Blood vessels, nervous regulation of the, 507
  pressure in, 485
  structure of, 480
Blood volume, 376
  shock and, 524
Body fat, 328, 630
  fluids, composition of, 632
  human composition of, 96
  image, 904
  water total, 634
Body temperature, 199
  clothes and, 209
  nervous control of, 205
  variations in, 207
  water, total, 629
Bohr effect, 573, 576
Bomb calorimeter, 152
Bone, 104, 929, 951
  composition of, 104
  conduction, 727
  formation and repair, 139
  growth in rickets, 144
  marrow, 390, 400
Botulinum toxin, 790
Bowman's capsule, 590
Bradycardia and oxygen, 560
  test, 123
Bradykinin, 220
Brain, blood flow, 820
  circulation in, 533
Brain, coronal section, 885
  diagram, 818
  membranes of, (fig.) *facing page* 664
  metabolism, 820
  parts of, 819
Brain stem, 847
  decerebrate rigidity, 852
  labyrinths and, 853
  reticular formation, 850
Bread, 97, 180, 181
Break 'shock,' 759
Breasts, 993
Breath-holding test, 550
Breath sounds, 547
Breathing, initiation of, 988
Breathing reserve, 550
Brewster's stereoscope, (fig.) 702
Brightness, range of, 688
British anti-lewisite, 74
Broca's area, 896
Brodmann's areas, 888
Bromsulphalein, 529
Bronchi, 538
Bronchial murmur, 547
  muscle, 546
Bronchoconstriction, 546
Bronchodilatation, 546
Brown-Séquard syndrome, 830
Brunner's glands, (fig.) *facing page* 263, 264
Brush border, (fig.) *near page* 593
Buffer nerves, 464
Buffers, 32–34, 621
Buffering power of the blood, 640
Bundle, A.V., 453, 473
Butter, 176
Buttermilk, 176

Cadaverine, 351
Cæcum, 276
Caffeine, 59
Caisson workers, 560
Calcarine cortex, (fig.) *facing page* 897
Calciferol, 143
Calcification, 105
Calcium, 99, 144
  absorption, 100
    and vitamin D, 100
  bones and teeth and, 101
  carbonate in flour, 193
  clotting of blood, 362
    milk, 174
  content of foods, 100
  daily allowance, (table) 190
  excretion, 101
  human blood serum, (table) 101
  in blood, 931
    cerebrospinal fluid, 100, 646
    foodstuffs, (table) 190
  metabolism and parathyroid glands, 929
  phosphorus ratio, 145
  requirements, 102
  serum, 929
Calculus, salivary, 222
Caloric stimulation of semi-circular canals, 860
Caloric tests, 861
Calorie, definition, 152

Calorie, intake and climate, 201
  requirements of man, 169
  value of one litre of expired air (nomogram), 162
Calorific value of oxygen, 154
Calorimeter, bomb, 152
  Atwater-Benedict, 158
  hand, 514
  oxy-, 153
Calorimetric measurement of blood flow, 514
Calorimetry, direct, 157
  indirect, 157
Canalis reuniens, 713
Canals, semicircular, 856
Cane sugar, 14
Cantharidin, 658, 809
Capacity of vascular system, 482
Capillaries and adrenaline, 519
  and tissue fluid, (fig.) 629
  ascorbic acid and, 141
  behaviour of, 517
  chemical control, 518
  contractile power, 518
  contraction of, dilatation and blood coagulation, (fig.) 369
  flow in, 506
  in unit area, 518
  nervous control, 517
  number, in guinea-pig muscle, 518
  pressure, 490
  pulmonary, 527
  skin colour and, 519
  structure, 480, 517
  vasomotion, 518
  vitamin C and, 141
  vitamin P and, 142
Capillary wall, effect of pressure on the passage of fluid through, (fig.) 636
Carbamino compounds, 581
Carbohydrate, antiketogenic substance, 332
  body fat and, 291, 328
  breakdown of: the Pasteur effect, 316
  calorie value of, 154
  in the diet, 194
  liver glycogen and, 291, 293
  metabolism, 288–322
    effect of anterior pituitary hormones on, 298, 300
    hormones of the adrenal cortex in, 305
    pancreas and, 301
  oxidation and insulin, 301
  protein-sparing action of, 189
  requirements, 194
Carbohydrates, 8–18
  absorption of, 269
  digestion of, 265
Carbon dioxide, acidæmia and, 641
  alkalæmia and, 641
  and blood pressure, (fig.) 515
  carriage of, 577
  combining power, 641
  dissociation curves, (fig.) 578
  in blood, 572
    carbamino compounds and, 581
    carbonic anhydrase and, 581
    estimation of, 572
    muscular exercise and, 794
    passage of gases through alveolar membrane, 582
    tension, arterial and venous, 572, 582
Carbon dioxide, in respiration, 555
    artificial, 569
    effects of increasing, (table) 555
    excess, in asphyxia, 514
    in hyperventilation, 515
  retention poisoning, 586
  vasodilator substance, 515
  vosomotor centre and, 514
Carbon monoxide, 404
  hæmoglobin, 405
  poisoning, 586
Carbonic anhydrase, 77, 240, 581
Carboxyhæmoglobin, 404, 405, 408
  absorption spectra, (fig.) *facing page* 406
Carboxylase, 77, 122, 314
Carboxylation, 131
Carboxypeptidases, 253, 265
Cardia, 225, 228, 231
Cardiac catheterization, 450
  cycle, 432
  denervation, (fig.) 461
  failure, 446, 528, 563
    œdema in, 620
  glands, 239
  impulse, 437
  index, 451
  muscle, absolutely refractory period and cardiac rate relation in dog, 429
  nerves, 463
Cardiac output and adrenaline, 517
  atrial pressure and, 445, 451
  exercise and, 800
  in anæmia, 586
  in heart-lung preparation, 445
  in man, 450
    dye method, 453
  peripheral resistance and, 499
  shock and, 524
  rate, 678
  reflexes, 462
  sphincter, 228
  tamponade, 437
Cardio-accelerator centre, 457
Cardio-aortic nerve, 463
Cardio-inhibitory centre, 462
Cardiogram and potassium of plasma, 475
Cardiometer, 433, 449
Carotene, 117
β-Carotene, 115
Carotenoids, 115
Carotid body, 466, 557, 561
  and respiration, 563
Carotid chemoreceptors, 513
Carotid loop, 615
Carotid sinus, (fig.) *facing page* 465, 513, 561, 564
  and blood pressure, (fig.) 513
  nerve, 464
Carotovagal compression, (fig.) 464
Carpopedal spasm, 929
Carr-Price reaction, 116
Cartilage, 106
Caseinogen, 173
  action of rennin on, 173
Castration, 950, 951
  androgens and œstrogens in urine after, 953
  cells, 970
Catabolic response to injury, 194
Catabolism, 2
  definition of, 2

Catalase, 70, 82, 98
Cataract, 673
Catechol, 81, 811
  amines, 941
    urinary output, 943
Catheterization, cardiac, 451
Catelectrotonus, 754
Cathepsin, 78
Cation exchangers, 38
Caudate nucleus, 884
Cell, definition, 1
  membrane, 737
  potassium, 737
  sap, 67, 78
  sodium, 737
Cell division, 1013
Cell of Schwann, 733
Cells of blood, 385
  of cardiac muscle, 425
  of Purkinje, 425, 866
  of Rouget, 517
  of Sertoli, 949
Cellulose, 17, 266
Central artery of retina, 682
  excitatory state, 834
  inhibitory state, 839
  reflex time, 832
Central sulcus, 883
  synaptic transmitter, 753
Centrifuge, human, 493
Centre, pneumotaxic, 554
  respiratory, (fig.) *facing page* 552, 553
  vasomotor, 507
Cereals, 178
  composition of products containing, (table) 179
Cerebellar connexions, 869
  tremor, 870
Cerebellum, (fig.) *near page* 833
  and decerebrate rigidity, 855, 873
  and gait, (fig.) 871
  and limb movement, (fig.) 874
  cerebellar ataxia, 870
  decerebellation, 869
  grey matter, 866
  histological appearance, 867
  injury, (fig.) 873
  intention tremor, 870
  lobes of, 866
  peduncles, 865
  removal of, 870
  scheme of connexions, 868
Cerebral circulation, 533
Cerebral cortex, 883
  cytoarchitectural fields of, (fig.) 888
  giant Betz cells, and function, 890
  grasp reflex, 896
  hæmorrhage, 896
  hemisphere, 883, 884, 886
    fibres, myelination of, 889
  histology, 887 (fig.) *facing page* 888
  motor areas, 890, 896
  motor homunculus, (fig.) 891
  occipital cortex, 898
  parietal lobes, 902
  premotor area, 895
  repsesentation of autonomic activities, 896
  sensory homunculus, 903
  structure, 887 (fig.) *facing page* 888
  temporal lobes, 901
Cerebrosides, 23
Cerebrospinal fluid, 644, 645
  blood plasma, chemical composition compared, (table) 646
  calcium, 100
  composition of, 646
  formation of, 646
  sample obtained, 645
Cervical ganglion, superior, investigations on, 746, 748
Cervix, 954
Cheese, 176
Cheilosis, 126
Chemical control of vascular tone, 514
Chemical score, 187
  senses, 661
Chemoreceptors, 466, 513, 563
Chemotaxis, 661
Chest, shape of, 540
  teleradiograms of, (fig.) *facing page* 424
Chewing, 218
Cheyne-Stokes respiration, 559
Chiasma, optic, (fig.) *facing page* 896
Children, calorie requirements, 168
  height, weight and surface area, 167
  underweight, 1021
Chitin, 17
Chloride and sweating, 205
Chloride shift, 578, (fig.) *facing page* 584
Chlorophyll, 413
Chocolate, 182
Cholagogues, 259
Cholanic acid, 26
Cholecalciferol, 143
Cholecystokinin, 259
Choleglobin, 414
Choleretics, 256, 259
Cholesterol, 24, 256
  in bile, 256
  in blood, 333
  metabolism of, 333
Cholic acid, 25, 355
Choline, 22, 135, 273, 327, 347, 353
  acetylase, 748, 753
  and fat absorption, 273
  deficiency, 327
  esterase, 74
  liver fat, 326
Cholinergic nerve, 753, 812
  nerves, diagram, 814
Cholinesterase, 788
Chondroitin sulphate, 18
Chorda tympani, 219, 223
Choreiform movements, 886
Chorionepithelioma, 979
Chorionic gonadotrophin CG excretion, 980
  gonadotrophins CG, 969, 975, 977, 979
  villi, 976
Choroid, 670
  plexuses, 645
Christmas factor, 365
  disease, 367
Chromaffin tissue, 941
Chromatic aberration, 677, 678
Chromatin, 65
Chromatolysis, 765
Chromophobes, 998
Chromoproteins, 54

Chromosomes, 65, 1013
  crossing over, 1017
  sets, 1013
  sex, 1015
Chronaxie, 756
Chvostek's sign, 930
Chylomicrogram, 273
Chylomicrons, 274
Chyme, 265
Chymotrypsin, 73, 253, 265
Ciliaris muscle, 674
Ciliary activity, 538
  ganglion, 901
Cingulate gyrus, 884
Circulation, 422
  cerebral, 533
  effect of acceleration, 492
    of gravity, 492
  fœtal, (fig.) *facing page* 985, 985
  in skin, 519
  peripheral, 480
  plan of, 423
  pulmonary, 527
  speed, 501
  through liver, 529
  time, 504
Circus movement, 477
Cirrhosis, 327
Cisternal puncture, 644, 645
Citrate and blood coagulation, 380
  and clotting, 362
Citric acid, 104, 317
  cycle, 316
Citrovorum, factor (CF) 132,
Citrulline, 342
Clasp-knife reaction, 843
Classification of the proteins, 53
Claude Bernard, 293, 307
Clearance tests, 604
    in man, 609
  urea, 607
Climacteric, 960
  hormonal basis of, 971
Climate, 212
  and calorie intake, 201
  clothes and, 209
  food eaten and, 201
  protein requirement and, 202
Clo, definition, 210
Clonus, 782
Clostridium botulinum, 790
Clot retraction, 366
Clothing, 209
Clotting, 174
  mechanism of blood, 361
  milk, 173
  time, 362
Coagulation and temperature, (fig.) 361
  factors in, (fig.) 362
  of blood, 361
  time, 362, 365
    capillary contraction, dilatation and, (fig.) 369
Coal gas poisoning, 586
Cobalamin, 134, 395
Cobalt, 107, 133
Co-carboxylase, 122
Cochlea, (fig.) 714
  electrophysiological aspects of, 717
  modern view of action of, 720
Cochlea, mechanical aspects of, 715
  position of, (fig.) 860
  theories of, 715
Cochlear microphonics, 717
  nucleus, 719
  nerve, central connexions, 721
Cocoa, 97
Cod-liver oil, 143
  (vitamin D), 142
Cœliac disease, 274
Coenzyme, A, 62, 77, 130, 316, 318, 329
  I, 62, 85, 127
  II, 62, 86, 127
Coenzymes, 76
  nucleotides, 60
Coitus, 950, 974, 975
Cold, effects on skin, 521
  in head, 712
  spots, 652
Colic, 279, 808
Collagen, 49, 53, 1024
  fibres, 651
Colloid, 34, 50
  osmotic pressure, 37, 193
Colloidal properties of proteins, 34
Colon, 276, (fig.) *facing page* 279
  absorption in, 279
  nerves, 277
  peristalsis in, 279
  X-ray photograph, (fig.) *facing page* 279
Colour-blindness, 691, 1019
  detection of, 692
Colour index, 406
  mixture, 690
  saturation, 689
Colour vision, 689
    theories of, 693
    with small fields, 693
Colours, complementary, 691
Coma, diabetic, 310
Comb growth test for androgens, 952
Comfort, 209
Commissural fibres, 888
Common chemical sensibility, 661
Competitive inhibition, 75, 134
Complementary colours, 691
Complete heart block, 454
Composition of the human body, 96
  of meals, 195
Conception, 974
Condensed milk, 176
Condensing enzyme, 317
Conditioned reflexes, 911
  analyser, 915
  and sleep, 907
  delayed, 914
  discrimination, 914
  external inhibition, 913
  extinction, 912
  in man, 915
  inhibition, 913
  internal inhibition, 914
  negative, 913
  reinforcement, 912
  simultaneous, 912
  specificity of, 913
  trace, 914
Conducting tissue of heart, 453
Conduction, air, 727
  bone, 727

Conduction, conditions influencing, 754
  deafness, 730
  in nerves, 737
  speed (Helmholtz), 746
  velocity in nerve, 743
Conduction rate, heart, 455
  saltatory, 740
Cone monochromats, 692
Cones, number of, 684
  of retina, 682
Conjugate movements of the eyes, 898
Conjugated proteins, 54
Conjugation and detoxication, 419
  of amino-acids, 347
Conjunctiva, 669
Consciousness, 864
Conservation of matter and energy, 3
Consonant, 707
  transitional, 709
  unvoiced, 707
  voiced, 707
Constant element, 326
Constipation, 279
Contact lens, 669, 680
Contraction, chemical basis of muscle, 776
Contrast, successive and simultaneous, 691
Convection, 202
Convergence, 675
Conversion factors, 1026
Convoluted tubules, 591
Cooking, effect of, 183
Copper, 77, 81, 87, 107
  and hæmopoiesis, 395
  sulphate in estimating specific gravity of blood, 375
Coprosterol, 25
Cord, spinal, 822–846
Corium, 651
Cornea, 669, 673
Corneo-sclerotic junction, (fig.) 674
Corona radiata, 888
Coronal plane, 819
Coronary arteries, (fig.) *facing page* 456
  circulation, 456
  disease, 476
  vessels, 426
Corpus callosum, 888
Corpus luteum, 956
  hormone, 963
  of pregnancy, 975
  after hypophysectomy in the rabbit, 968
  test for luteal activity, action on endometrium of rabbit, (fig.) 964
Corpus striatum, 884
Corpuscle in alveolar capillary, 577
Corpuscles of Hassall, 944
Corpuscular or globular proteins, 48
Cortex, adrenal, 936
  and autonomic system, 896
  auditory, 901
  cytoarchitectural fields of human, (fig.) 888
  frontal, 896
  motor, 890
  occipital, (fig.) *facing page* 897
  parietal, 902
  premotor area, 895
  sensory, 897, 902
  structure of, 887
  temporal, 901
  visual, (fig.) *facing page* 897
Corti, organ of, 714
Cortical insufficiency, graph, 940
  retina, (fig.) *facing page* 897
Corticosterone, 28, 937
Cortisone, 937
Cough syncope, 436, 523
Coughing, 564
Cozymase, 62, 85
Cramp, stoker's, 205
Craniosacral outflow, 804
Cream, 176
Creatine, 351
  phosphate, 61, 240
  tolerance, test, 353
Creatinine, 351, 599
  in measurement of glomerular filtrate, 605
Creatinuria, 353
Cretinism, 924
Cretins, (fig.) *following page* 924
Cricopharyngeus, 226, 708
Crista basilaris, 714
  dividens, 985
Cristæ ampullares, 856
Critical frequency of flicker, 697
Crossed extension reflex, 837
Crossing over of chromosomes, 1017
Crotonase, 329
Crushing of muscle, 774
Cryptorchidism, 948
Crypts of Lieberkühn, 264
Cumulus ovaricus, 956
Cupola, 858
Curare, 749, 790, 930
Curds, 241
Current, alternating, 757
  faradic, 757
  galvanic, 757
  high frequency, 758
  radio frequency, 757
  of injury, 743
Curve, CO_2 dissociation, 579
Cushing's syndrome, 1008
C_2 units, 299, 329
Cutaneous sensations, anatomical basis, 659
  spider nævi, 965
Cyanhæmatin, 406
Cyanide, 84, 762
  and the cytochrome system, 84
Cyanocobalamin, 132, 395
Cyanosis, 587
Cyclopenteno-phenanthrene ring, 25
Cysteine, 354
Cystic duct, 251
Cystine, 327, 354
Cystinuria, 284, 355
Cytidine, 59
Cytochemistry, 66
Cytochrome, 83, 98
  c reductase, 88
  oxidase, 67, 78, 82, 84
Cytoplasm, 64
Cytosine, 57, 65
Cytotrophoblast, 969, 976

Dairy produce, composition of, (table) 176
Dalmation coach hound, 356
Dark adaptation, 696
  flicker, critical frequency of, 697
  vitamin A deficiency, 118, 698

Dartos muscle, 948
Deacylase, 330
Dead space, 549
Deaf aid amplification curve, 729
Deaf aids, 725
Deafness, 726, 728, 732
  conduction, 730
  high frequency, 725
  in children, 726
  measurement of, 724, 728
  perceptive, 730
  traumatic, 727
Deamination, 339, 340
Deaths, fœtal, 1016
Decamethonium, 790
Decarboxylation of amino-acids, 350
Decerebellate dog, 870
Decerebrate man, (fig.) *facing page* 864
  rigidity, 852, 873
Decerebration by arterial ligation, 855
  in primates, 863
Decholin in estimating circulation time, 504
Decibel notation, 710
Decidua, 972, 976
Decidual cells, 976
  reaction, 963
Decompression sickness, 560
Decorticate animal, 878
  preparation, 854
  rigidity, 863
Decussations, 847
Defæcation, 277, 830
  nerve impulses in, (fig.) 277
Defibrinated blood, 362
Degeneration in spinal cord, 825
  in peripheral nerve, 822, 763
  transneuronal, 824
Deglutition, 225–229
Dehydration, 635
L-dehydroascorbic acid, 137
7-dehydrocholesterol, 25, 143
Dehydrogenases, 69, 80, 84
Dehydroisoandrosterone (formula), 951
Delayed conditioned reflex, 914
Demarcation potential, 743
d'emblée reflex, 836
Denaturation of protein, 52
Dental decay, 145
Dendrites, 733
Dentate nucleus, 867
Deoxycholic acid, 25
Deoxycorticosterone acetate (DCA), 940
Deoxyribonuclease, 65, 66
Deoxyribonucleic acid (DNA), 57, 418
Deoxyribonucleoprotein, 65
Deoxyribose, 57
Depolarization, 739, 750
Depressor nerve, 461
Depth of field, 677
  perception of, 702
Dermatome, 828
Descending degeneration in cord, (fig.) 825
  paths in spinal cord, 825
Desquamation, 522
Detoxication, 419
Detrusor muscle, 624
Deuteranopia, 692, 693
Deuterium, 286, 325
  oxide, 630
Dextran, 594
Dextrins, 17
Dextrose, 10
Diabetes insipidus, 613, 619, 879, 932, 1011
  effect of neurohypophysial extracts, (fig.) 1011
  removal of posterior lobe of pituitary and, 879, 1011
Diabetes mellitus, 215, 284, 307, 330, 932
  and growth hormone, 300
    hunger, 215
    piqûre, 307
  blood cholesterol in, 333
  blood-sugar curves from patients with, 309, 310
  effect of insulin on, (fig.) 309
  hypophysectomy in dogs and, 300
  induced in animals by administration of pituitary gland extracts, 300
  ketosis in, 310
  treatment, 311
Diabetic coma, 310
  ketosis, 642
  mothers and large children, 1021
Diabetogenic effect of adenohypophysial extracts, 300, 1006
Dialysis, 36
Diaphragm, 540
  pain related to, 808
Diarrhœa, 279
Diastase, 294
  in urine, 253
Diastolic blood pressure, 485
Diathermy, 758
Dichromats, 692
Dicoumarin, 362, 364
Diencephalon, 876
Dicrotic wave, 433
Diet, alipotropic, 327
  body fat and, 328
  carbohydrate requirements, 194
  fat requirements, 194
  insulin sensitivity and, 303
  life-span in rats and, 1022
  meals, analysis of, 195
  normal, 184
  protein requirements, 184
Dietary allowances, recommendations of Committee on Nutrition of B.M.A., 190
Diethylstilbœstrol, 965
Difference tones, 712
Differential centrifugation, 66
  count of white cells, 399
Diffusion capacity of alveolar membrane, 584
Digestion in stomach, 245
  in the small intestine, 265
  of carbohydrate, 265
    cellulose, 266
    fats, 267
    nucleic acids, 267
    proteins, 265
  salivary, 222
Digestive enzymes, (table) 266
Dihydroxy-acetone, 8
Dihydroxyphenylalanine (DOPA), 138, 349
2: 5-dihydroxyphenylpyruvic acid, 348
Diiodotyrosine (formula), 920
3: 5-diisoprophylfluorophosphonate, 74
2: 3-dimercaptopropanol, 74
Dinitrofluorobenzene, 32

2: 4-dinitrophenol, 319, 762
Dinucleotide, 63
Diodone, 600, 605
  clearance, 606
  Tm, 606
Diodrast, 600, 605
Diœstrus, 960
Dioptre, definition, 673
Dipeptidases, 265
Dipeptide, 44
Diphasic records, 743
1: 6-Diphosphofructose, 312
1: 3-Diphosphoglyceric acid, 312
Diphosphopyridine nucleotide (DPN), 62, 85
Diphosphothiamine, 314
Diploid number of chromosomes, 1014
Diplopia, 700
Direct calorimetry, 157
  oxidative pathway, 316
Disaccharide, 8, 13–14
Discrimination, 656
Distance, perception of, 702
Dissociation constant, 30, 641
Diuresis, 613, 941
  osmotic, 600
  pressure, 596
  water, 613
    inhibition of, by a rise of osmotic pressure in the carotid artery, (fig.) 614
Diuretics, 600
Diver, 560
Division, cell, 1013
Dizziness, 902
DNA, 57, 63, 65
D/N ratio, 307
Dominant character, 1016
Donnan equation, 737
  equilibrium, 36
DOPA, 348
Dorsomedial nucleus of thalamus, 881
Double vision, 700
Douglas bag, (fig.) 161
DPN, 62, 85, 127, 313, 314, 316, 332
  (diphosphopyridine nucleotide),
Draught, 994
Drugs, blocking transmission, 748
Ductus arteriosus, 985, 988
  cochlearis, 713
  venosus, 985, 988
Duodenal bulb, (fig.) *facing page* 260
  glands, (fig.) *facing page* 263, 264
  juice, 254
Duodenum, 260
  section through, (fig.) *facing page* 260
Duplex theory, 685
Dura mater, 644
Dwarfism, 1004, 1008
Dye method for blood volume estimation, 376
  for cardiac output, 453
Dyes, anticoagulant, 363
Dysmetria, 875
Dyspnœa, 556, 562

Ear, basilar membrane, 714, 720
  characteristic curve, 713
  drum, 711
  external and middle, 711
  harmonic distortion and, 712
  internal, 714
Ear, summation and difference tones, 712
Echo, 723
Eck fistula, 284
Ectopic foci, 473
Edinger-Westphal nucleus, 675, 677, 899
Effective filling pressure, 436
  filtration pressure, 597
Effects of starvation, 195
Efferent fibres, 769
  nerves, definition, 5
  γ efferents, 841, 854
Efficiency, 198
  gross, 165
  net, 165
  of bicycle-pedalling, 784
  of heart, 448
  of human body, 164
  of kidney, 600
Effort syncope, 523
Egg, 177
Egg-white injury, 131
Einthoven's string galvanometer, 468
Ejaculation, 950
Elasticity of aorta, (fig.) 481
Electrical units, 1027
Electrocardiogram, 468, 472, (fig.) 474
  clinical applications, 473
  diagram showing beginning and end of various parts, 472
  in coronary disease, 476
  jugular pulse and, (fig.) 499
  leads used in clinical electrocardiography, 471
  theoretical derivation, 469
Electrocortin, 28, 938
Electrodes, 754
Electroencephalogram, 904
  and flicker, 906
  in hypoglycæmia, 907
  in sleep, 907
Electrogram, (fig.) 476
Electrolytes, movement of, through capillaries, 636
Electromyograms, (fig.) 792
Electromyography, 792
Electron micrograph of toad muscle, (fig.) *facing page* 772
  of intestine, (fig.) *facing page* 260
  of liver cell, (fig.) *facing page* 68
Electrophoresis, 50, (fig.) 56, 370
Electrophrenic respiration, 570
Electroretinogram, 694
Embryo, 179
Emetic centre, 249
Emetics, 249
Emmetropic eye, (fig.) 679
Emotion and gastric emptying time, 234
  and lactation, 994
  urinary output, 610, 619
Emotional glycosuria, 306
  stress, 1002
Emphysema, 546
Emptying time of stomach, 234
Emulsification in intestine, 272
Encephalitis lethargica, 909
Endemic goitre, 927
End-groups, 46
Endocrine glands, 918
Endogenous nitrogen metabolism, 340
Endolymph, 714, 856

Endolymphatic potential, 717
Endometrium, 955
  in menstrual cycle, 957, (fig.) *facing page* 960
  in pregnancy, 972, 976
Endoneurium, 734
Endoplasmic reticulum, 67
Endosperm, 178
End-plate delay, 751
  potentials, miniature, 789
Endurance or 40 millimetre test, 550
Enema, 279
  barium, (fig.) *facing page* 279
Energy, daily needs, (table) 171
Energy exchange, 152–172
Energy expenditure by different species, (table) 166
Energy output, methods for estimating, 157
  requirements, 170
Energy-rich phosphate bonds, 60, 313, 318, 776
Enolase, 313
Enopthalmos, 677
Entero-gastric reflex, 235
Enterogastrone, 235
Enterohepatic circulation, 256
Enterokinase, 76, 252, 265
Environmental conditions, 212
Enzymes, 69
  activators, 75
  anti-, 75
  apoenzymes, 77
  coenzymes, 77
  competitive inhibitors, 74
  digestive, (table) 266
  inhibitors, 73
  intestinal, 265
  intracellular and extracellular, 69
  molecular weights of, 70
  of gastric juice, 241
  optimum pH, 71
  optimum temperature, 70
  prosthetic group in, 76
  specificity of, 73
  structure of, 76
  zymogens, 76
Eosinophil polymorph, 397
Eosinophilia, 400
Ephedrine, 812
Epidermis, 651
Epididymis, 949
Epiglottis, 225
Epileptic fit, 892
Epileptic seizures and motor areas of cerebral cortex, 892
Epinephrin, 941
Epineurium, 737
Epithelium intestinal, 268, (fig.) *facing page* 272
Equal energy luminosity curve, 686
Equivalent, 29
Erepsin, 265
Ergocalciferol, 143
Ergometer, 164
Ergosterol, 25, (formula) 143
Eructation, 227
Erythema ab igne, 522
  of skin, 522
Erythrocyte, in childhood, 402
  life of, 390
  -maturing factor, 394
Erythrocyte, sedimentation rate, 389
Erythrocytes, 385
  survival of transfused, 390
Erythropoiesis, 390
Eserine, 74, 221, 233, 678, 789, 813
Essential amino-acids, 351
  fatty acids, 194, 325
Ethanolamine, 23
Ethereal sulphate, 419
Ethinyl œstradiol, 963
  testosterone, 965
Ethylene, reductase, 329
Eunuch, 951
Evans blue, 376
Eve's method of artificial respiration, 569
Exchangers, ion, 38
Excitation at motoneurones, 751
Excretory capacity, tubular, 606
Exercise, 583, 793
  and blood lactic acid, 777, 799
  and fatigue, 787
  and muscle temperature, 797
    blood pyruvate, (fig.) 777
    respiration, 565
  cardiac output, 800
  chemical changes, 796
  circulatory and respiratory effects, 793
  lactic acid, 516
  muscular, 793
  oxygen consumption, 800
  R.Q. in, 795
  vasodilatation, 516
Exogenous nitrogen metabolism, 340
Exophthalmos, 925, 1003
Expiration, 540
Expiratory effort, maximum, 550
  reserve volume, 545, 547
Expired air, 551
  caloric value, 162
Extension reflexes, 837
External ear, 711
  inhibition, 913
Extinction of conditioned reflex, 912
Extracellular fluid, 629, 632
  formation of, 636
  hydrogen-ion concentration of, 640, 643
Extractives, 178
Extramedullary hæmopoiesis, 389
Extrasystoles, 430, 473, 474
Extrinsic factor, 394
Eye, accommodation, 674
  autonomic fibres to, 901
  focal length, 673
  fundus of, (fig.) *facing page* 680
  optics of, 672
  protective mechanisms, 670
  righting reflexes, 863
  structure of, 669
Eyes, conjugate movements of, 700, 898
  diplopia and, 700
  fixation, 700
  motor area of cerebral cortex and, 895
  movements of, 700
  nystagmus, 700, 701, 861, 870, 875
  occipital cortex and, 898
  oculo-sympathetic fibres, diagram of course, 901
  saccadic movement of, 700
  visual reflexes, 900
Eyelids, 670

Facilitation, 747
  of monosynaptic reflex, 750
Factor V and clotting, 365
Factor VII, 364
Facultative reabsorption, 598
FAD, flavine-adenine dinucleotide, 62, 87
Fæces, 281
Failure, circulatory, 524
Fainting, 523
  due to fall in cerebral blood pressure, 533
Famine, œdema, 373
Faradic current, 757
  stimulation, 759
Fasting, 214
Fat, 19
  absorption of, 270
  and glycogen, 297
  calorie value of, 154
  content of the diet, 194
  conversion to carbohydrate, 297
  digestion of, 267
  droplets (chylomicrons), 273
  emulsion, 272
  essential fatty acids, 194
  fatty acids labelled with deuterium, 325
  from carbohydrate, 328
    protein, 328
  gastric inhibitor, 245
  depots, 324
  in blood after meal, 323
  iodine value, 21
  milk, 173, 996
  lipid concentration of plasma, 324
  liver function and, 325
  metabolism, 323–334
    diagrammatic representation, (fig.) 325
    in diabetes, 308
  of tissues, 324
  oxidation, 328, 418
  partition hypothesis of absorption, 271
  protein-sparing action of, 189
  requirements, 194
  saponification value, 20
  stores, 324
  transport by placenta, 982
Fatigue, 783, 836
  of muscle, 786
    nerve, 760
Fatty degeneration, 326
'Feed-back' mechanism, 1001
Feeling, threshold of, 724
Fehling's solution, 11
Female genital organs, 954
  sex hormones, 27, 961
Fenestra vestibuli, 713
  cochleæ, 713
Fermentation, 11, 69
Ferritin, 98, 392
Fertility, 972
  ovulation and, 972
Fertilization, 1015
  of ovum, 974
Fever and salivation, 222
Fibrillation, 790, 793
  atrial, 473, 475, 477
Fibrin, 362
Fibrinogen, 362, 364, 369, 370
Fibrinolysis, 367
Fibrous proteins, 48
Ficin, 76
Fick principle, 449, 529, 533
  in kidney, 607
  in mammary gland, 996
Field of vision, 687, 688, (fig.) *facing page* 897
Filtration angle, 672
  fraction, 596, 607
  pressure, 597
Final common path, 835
Fish, 178
Fistula, gastric, 235
  Thiry-Vella, 267
Fixation, 700
  of carbon dioxide, 131
Fixity of internal environment, 198
Flaccid paralysis, 893
Flare, 520
Flatus, 279
Flavine-adenine dinucleotide (FAD), 62, 87
Flavoproteins, 86, 126
Flexion and extension reflexes, characteristics of, 837
  reflexes, 750, 830
Flicker, 697
  and e.e.g., 906
  photometry, 691
Flocculonodular lobe, 866
Flour, 103, 180
Flow, blood, 501
  in vessels, 500, 506
Fluids of body, 629
Fluorescein and circulation time, 504
Fluoride and clotting, 363
Fluorine, 107
  in bones and teeth, 107
Flutter, atrial, 473, 475, 477
Flux of ions, 738
Fœtal circulation, (fig.) *facing page* 985
  hæmoglobin, 403
  respiration, 551
Fœtus, 984
  birth weight and size, 984
  composition of human, 984
Folic acid, 131
  oxygen saturation of blood, 985
Folinic acid, 132
Follicular hyperkeratosis, (fig.) 118
  phase, 957
Follicle-stimulating hormone (FSH), 968
Food, 173–197
Foods, ascorbic acid in, 141
  calcium in, 100
  effect of cooking, 183
  iron in, 97
  nicotinic acid in, 129
  phosphorus in, 100
  phytic acid in, 102
  riboflavin in, 127
  thiamine in, 124
  tocopherol (vitamin E) in, 148
  vitamin A in, 120
  vitamin B_{12} in, 133
  vitamin D in, 146
  vitamin K in, 150
Foot, venous pressure, 489
Foramen ovale, 985
Forced grasping, 896
Formant, 709
Formed elements of blood, 385
Formol titration, 43
Fortification, 180

- Fovea centralis, 682
- Fractional gastric analysis, 246
- Fractionation, 835
- Frequency distortion, 725
  - distribution curves of diameters of red cells, (fig.) *facing page* 385
- Friedman test, 979
- Frog's heart, 427
- Frontal cortex, 897
  - association areas, 897
  - leucotomy, 897
- Frost-bite, 522
- Fructose, 12, 288, 296, 305
  - and liver glycogen, 293
  - in liver disease, 296
- Fructosuria, 305
- Fruits, 181
- FSH (follicle-stimulating hormone), 968
- Functional residual capacity, 547
- Fundus of eye, (fig.) *facing page* 680
  - stomach, (fig.) *facing page* 231, 232
- Furanose, 10

- Gait in cerebellar disease, 871
  - in hemiplegia, 894
- Galactose, 12, 288, 296, 305
- Galactosides, 12, 14
- Galactosuria, 305
- Gall-bladder, (fig.) *facing page* 255, 257
  - and triple intubation, 258
  - bile, composition of, (table) 255
  - movements of, 257
  - pain, 808
  - volume of, 257, 259
  - X-ray photograph, (fig.) *facing page* 255
- Gall-stones, 257
- Galvanic current, 757
  - stimulation, 758
- Galvanometer, string, 468
- Gametes, 1013
- Gamma efferents, 772, 841, 854
- Gammexane, 136
- Ganglion, blocking drugs, 748
  - posterior root, 823
  - potentials, (fig.) 747
  - spinal, 823, (fig.) *facing page* 824
  - superior cervical, 746
  - sympathetic, 746
- Gas analysis, 552
  - intestinal, 280
- Gases and alveolar membrane, 582
  - in blood, 572
  - in water, 571
- Gastrectomy, 214
- Gastric analysis, 246
- Gastric contractions, adrenaline and, 234
    - and hunger, 214
    - inhibited by water, (fig.) 233
  - digestion, 245
    - glands, 236
    - inhibitors, 235, 245
  - emptying time, 234
  - fistula, 233, 235
  - glands, 238
    - hydrochloric acid, 239
      - formation of, 240
  - inhibitors, 245
- Gastric juice, collection of, 235
    - composition of, 237, 394
    - enzymes of, 241
    - histamine and, 238, 242, 247
    - hydrochloric acid, 239
    - insulin test and, 248
    - pH of, 246
    - secretion of, 235
  - lipase, 237
  - motility, 231 (fig.) 233
    - adrenaline and, 234
    - atropine and, 233
    - eserine and, 233
    - food and, 232
  - mucosa, (fig.) 236, 394
    - resistance to enzymes, 241
  - pits, 235, 236
  - secretion, 241
    - cephalic or reflex phase, 244
    - chemical mechanism of, 242
    - enzymes of, 241
    - gastric or humoral phase, 244
    - gastrin and, 242
    - inhibition of, 235, 245
    - insulin and, 248
    - nervous mechanism of, 241
  - tetany, 932
- Gastrin, 242, 945
- Gastrocolic reflex of Macewen, 279
- Gastroscope, 233
- Geiger counter, 287, 922
- Gelatin, 46, 53, 187
- Genes, 1016
  - blood groups and, 1017
  - enzymes and, 1020
- General adaptation syndrome, 939
- Genesis of tetanus, (fig.) *near page* 784
- Genomes, 1013
- Genotype, 1017, 1018
- Germinal epithelium, 949, 956
- Gestation, period of, 990
- Germ, 179
- Giant cells in the lungs, 400
- Gigantism, 1007
- Girl in puberty, 956
- Glands, apocrine, 652
  - ceruminous, 711
  - duodenal, 263, 264
- Glaucoma, 671, 680
- Gliadins, 53
- Globin, 54, 404
- Globular or corpuscular proteins, 49
- Globulins, 53, 369, 371
- γ-globulins, 371
- Globus pallidus, 884
- Glomerular arterioles, 596
  - capsule, 590, (fig.) *near page* 593
  - filtrate, 594, 603
  - filtration fraction, 596, 607
  - filtration rate, 604
  - filtration rates in new-born, 623
  - fluid, 594
  - function, 593, 595
  - pressure, 595, 597
- Glomerulus, 590, (fig.) *following page* 592
- Glottis, 705
- Glucagon, 304, 935
- Glucocorticoids, 938
- Glucogenetic amino-acids, 345
- Gluconeogenesis, 296, 418

Gluconic acid, 12
Glucosamine, 12
Glucose, 9, 10
  conversion to fat, 291, 328
  estimation of, 288
  formed from fat, 297
  kidney clearance and, (fig.) 608
  oxidation of, 292
  structure of, 9
  tolerance test, 288
  tubular function and, 605
Glucosides, 11, 12
Glucuronic acid, 12, 419
Glucuronides, 419
Glutamic acid, 820
Glutaminase, 622
Glutamine, 346, 622
Glutathione, 45, 89, 312, 355, 673
Glutelins, 53
Gluten, wheat, 274
Glycerol, 19
Glycine, 40, 46, 185, 352
Glycocholic acid, 255
Glycogen, 16, 418, 455
  and adrenal cortical hormones, 305
  disease, 284, 311
  in liver, 291, 293, 297
Glycogen in breakdown, 293
  fat as source of, 297
  formation of, 293
  from non-carbohydrate precursors, 296
  fructose and, 296
  in muscle, 291
  stores of the body, 297
  synthesis, 294
Glycogenesis, 291, 295
  insulin and, 296
Glycogenic amino-acids, 345
  function of the liver, 293
Glycogenolysis, 291, 293
Glycolysis, 294, 311, 314, 316, 320
Glycoproteins, 18, 54
Glycoside, 12
Glycosuria, 300, 305, 1005
G/N ratio, 307
Goitre, 106, 925
  endemic, 927
Goitrogens, 925
Golgi tendon organ, 771
Gonadotrophic hormones, 968, 1002, 1005
  chorionic, 969, 980
  excretion of, 969
  excreted in pregnancy, (fig.) 980
  menopause and, 969, 972
  puberty and, 969
Gonads, and puberty, 969, 971
Gout, 357
Graafian follicle, 956
Grain, wheat, 180
Gramophone audiometer, 728
Granulocyte series, 396
Grasping, forced, 896
Grassmann's laws, 690
Graves' disease, 925, 1003
Gravity and circulation, 492
Gray ramus communicans, 805
Growth, 1021
  effect of lysine and tryptophan, (fig.) *facing page* 185
  factors, 1021
Growth, hormone, 216, 301, 1003, 1005
Guaiac test for blood, 83
Guanase, 355
Guanidine, 352
Guanido-acetic acid, 352
Guanine, 57, 65
Guanosine, 57
Guanylic acid, 60
Gums, 17
Gustation, 664
Gustatory sweating, 668
Gyrus, (fig.) 884

'H' substance, 521
Hæm, 404, 409
Hæmagglutinins, 380, 383
Hæmagglutinogens, 380, 383
Hæmatin, 409
Hæmatocrit, 377, 981
  readings, 377
Hæmatoporphyrin (formula), 412
Hæmin, 83, 406, 407
Hæmochromogen, 407, 410
  absorption spectra, (fig.) *facing page* 406
Hæmoconcentration, 524
Hæmodilution, 524
Hæmoglobin, 83, 98, 404, 575
  absorption spectra, (fig.) *facing page* 406
  breakdown of, (fig.) 413, 414
    products, fate of, (fig.) 414
  buffering action, 580
  cyanosis and, 587
  estimation of, 405
  fœtal, 403
  gas exchange, (fig.) 584
  level, infancy, 401
  mean corpuscular concentration (MCHC), 392, 407
  molecular weight, 405
  pigments related to, 408
  to oxyhæmoglobin, reaction time, 577
Hæmoglobinometers, 405
Hæmolysins, 379
Hæmolysis, 387
Hæmolytic jaundice, 416
  reaction, 381
Hæmophilia, 366, 1019, 1020
Hæmopoiesis, 389
  and ascorbic acid, 395
  copper in, 395
  thyroid gland, 395
  nutritional factors in, 392
  protein in, 395
Hæmorrhage, effects of, 524
Hæmosiderin, 98, 392
Hæmostasis, 366, 368
Hair, 951
Hairs as sensory receptors, 655
Haldane hæmoglobinometer, 405
Haldane's blood gas-analysis apparatus, (fig.) 572
  method of estimating hæmoglobin, 405
Half-life, of body protein, 339
  of radioactive isotopes, 287
Hamburger phenomenon, 579
Haploid number of chromosomes, 1014
Harmonic distortion, 712
Harvey's work, 422

Hearing, 705
  acuity, 724
  area of useful, 724
  frequency distortion, 712
  loss, 726
    in cortical lesions, 902
  masking, 728, 902
  method of measuring intelligibility, 725
  Rinne's and Weber's tests, 730
  threshold, 724
Heart, 424
  acceleration, 466
  anatomy, 424
  apex beat, 438
  block, 454, 476
  cardiac impulse, 437
  cardiac output in man, 450
  conduction through, 453
  coronary circulation, 456
  denervation, 461
  diagram of nerve supply, 458
  disease, pain, 808
  effect of vagus on, (fig.) 460
  efficiency of, 448
  embryology of, 424
  failure, 446, 563, 620
  flutter and fibrillation, 477
  frog, 427
    acetylcholine, perfusion with, 459
    'all-or-nothing' law, 428
    pacemaker (sinus), 428
    perfusion of, Ringer's solution in, 430
    refractory period, 428
  histology of, 425
  law of the heart, 446
  -lung preparation, 440
    -kidney perfusion, 614
  mammalian, activities of the, 431
  metabolism of, 455
  movements, 437
  muscle-cells, (fig.) *facing page* 424
  nerve supply, (fig.) 458
  nervous regulation of, 457
    carotid body and, 466
    factors causing acceleration and slowing, 466
    right atrial reflex, 466
  output, 449
    and atrial pressure, 451
  oxygen consumption, 449
  perfusion of, 430
    mammalian, 431
  pressure in, 433
  rate, 467
    after atropine, 459
    age and, 467
    and temperature, 468
    denervation, (fig.) 461
    depressor nerve, 460, 463
    inflation of lungs and, 496
    respiration and, (fig.) 495
    vagal section, (table) 459
    vagus on, (fig.) 460
  refractory period, 428
  return of blood to, 436, 442
  sequence of events in the cardiac cycle, 432
  size of, 426
  slowing, 466
  sounds, 438
Heart, sympathetic nerves, 462
  systolic output, 449
  valves, 425
  venous return, 436
  weight of, 426
  work of, 447
Heartburn, 229
Heart-lung preparation, 440, 441
  arterial pressure, 442
  cardiac metabolism, 455
    denervated heart, 448
    innervated heart, 448
  Starling's law and, 446
    cardiac failure, 446
  venous inflow, 442
Heat balance, 198
  erythema, 522
  loss, 203
    and sweat, 203
    in insensible perspiration, 203
  production and heat loss, balance between, (fig.) 198
    in muscle, 781, 790
  in nerve, 761
  of combustion, 152
  stagnation, 213
Heavy hydrogen, 286
  water, 286, 630
    and total body water, 630
Height, maximum, 1021
Helicotrema, 713
α-helix, 49
Hemianæsthesia, 881
Hemianopia, 899
Hemiballismus, 886
Hemiplegia, 893
Hemisection of spinal cord, 830
Henderson-Hasselbalch equation, 641
Heparin, 18, 362, 363, 945
Hepatectomy, 284
Hepatic artery, 529
  function, index of, 257
    test (fructose), 296
    tests, 421
Heredity, 1013
  phenomena of, 1016
Hering-Breuer reflex, 562
Herring bodies, 999
Herpes zoster, 512
Heterozygous individuals, 1016
Hexamethonium, 749
Hexœstrol, 965
Hexokinase, 295, 301
  and pituitary hormones, 296
Hexonic acid, 12
Hexosamine, 12
Hexuronic acid, 12
Hiccough, 564
High altitude, 589
  effect of, 557
    oxygen pressures at, (fig.) 558
  -frequency deafness, 725
  -currents, 758
'High-temperature short-time' pasteurization, 175
Hippocampus, 876
Hippuric acid, 347, 419
  index of liver function, 347, 419
Hippus, 678
Hirudin, 362, 364

Histaminase, 350
Histamine, 238, 242, 350, 400, 521, 525, 546, 945
  and bronchial muscle, 546
  and gastric juice, 247
  and pain, 658
  causing bronchoconstriction in guinea-pig, 546
  flare, 521
  gastric juice and, (fig.) 247
  in intestine, 945
  shock symptoms in animal induced by, 525
  skin and, 521
Histidine, 350, 945
Histiocytes, 398
Histones, 54, 65
Histotoxic anoxia, 587
Holder process of pasteurization, 175
Holger Nielson method of artificial respiration, 567
Homocysteine, 354
Homogentisic acid, 349
Homoiostasis, 3, 198, 620
  and blood-sugar, 299
Homoiothermic animals, 198
Homozygous individuals, 1016
Homunculus, motor, 891
Honey, 182
Hopping reaction, 863
Hormones, 9, 918–947
  local, 945
  pituitary, 998–1012
  reproductive, 948–997
Horner's syndrome, 677
Horse-power, 165
Houssay dog, 300
H-substance, 521
Human body, composition of, 96
  electrocardiography, 471
  milk, composition, 97, 173
  placenta, (fig.) 976
Humidity, 212
Humoral theory, 810
Hunger, 214
  and blood sugar, 215
    satisfaction, 215
  gastric contractions, 232
  -pains, 297
Hyaluronic acid, 17
Hyaluronidase, 17, 975
  of semen, 950
Hydatiform mole, 979
Hydrocephalus, 646
Hydrochloric acid in stomach, 239, 246
  production by oxyntic cell, (fig.) 239
Hydrogen acceptor, 81
Hydrogen bond, 49, 65
  -ion concentration, 30
    and kidney, 620
    of the blood, 641
  peroxide, 82
Hydrotropic power, 256
β-hydroxybutyric acid, 332
    dehydrogenase, 332
5-hydroxyindoleacetic acid, 946
5-hydroxytryptamine, 349, 400, 946
  and pain, 658
Hyperæmia, reactive, 515
Hyperæsthesia, 830
Hyperalgesia, 659
Hyperglycæmia, 288, 943, 1005
  pancreatic islet tissue, damage to, 932
Hyperglycæmic-glycogenolytic factor, 304, 935
Hyperinsulinism, 935
Hyperkeratosis, (fig.) *facing page* 121
Hypermetropia, 679, 680
Hyperparathyroidism, 931
  and bones, (fig.) *near page* 932
Hyperpnœa, 555
Hyperpolarization, 750
Hypersomnia, 879, 909
Hypertensin, 612
Hypertension, 611
  experimental, (fig.) 611
  in man, 612
Hyperthyroidism, 920, 925, 1003
  basal metabolic rate (B.M.R.) in, 924
  effect of iodine, 926
  goitrogens and, 925, 927
  muscular tremor before and after thyroidectomy, (fig.) 926
  thiourea and thiouracil in, 927
  thyrotrophin and, 923, 1003
Hypertonic solution, 37
  solutions and urinary output, 615
Hyperventilation, 555
  tetany, 931
Hypervolæmia, 620
Hypofunction of adenohypophysis, 1004
Hypogastric nerves, 975
  plexus, 277
Hypoglycæmia, 311, 907, 935, 943
  electroencephalogram and, 907
  insulin overdose and, 311
  liver disease, 297
  symptoms of, 302
Hypoparathyroidism, 929
Hypophysial portal vessels, 1001
Hypophysectomy, 923
  in pregnancy, 977
Hypophysis cerebri, 998
  blood supply, 1000
  nerve supply, 999
Hypopituitarism, 1002, 1008
Hypoproteinæmia, 193, 373
Hypothalamic nuclei, (fig.) 877
Hypothalamo-hypophysial tract, 877, 1000
Hypothalamus, 876
  afferent nerve pathways, 876
  anæsthetics on, 879
  and gastric motility, 235
  and hunger, 215
  and œstrus, 970
  antidiuretic hormone and, 879
  body temperature and, 206, 878
  efferent nerve pathways, 876
  obesity, 879
  parasympathetic effect, 878
  pituitary, anterior lobe and, 1000
  sham rage and, 878
  sleep, prolonged, and, 879
  stimulation of, 878
  sympathetic effects, 878
Hypothyroidism, 920, 924
Hypotonia, 870
Hypotonic solution, 37
Hypoxanthine, 59, 355
Hypoxia, 559

^{131}I, 922
Icteric index, 416
Identical twins, 1017
Idioventricular rhythm, 476
Ileocolic sphincter, 262
Ileum, 260
  lymph follicle, (fig.) *facing page* 260
  segmentation, 261
Inborn errors of metabolism, 284
Incisura angularis (fig.) *facing page* 232
Incontinence, 627
Incus, 711
Indican, 349, 420
Indirect calorimetry, 157
Indole, 281, 349
Indoxyl, 349, 420
Inertia flywheel, 785
Infancy, amino-nitrogen absorption in, 335
  and childhood, physiological changes in blood during, 402
  antidiuretic hormone in, 623
  basal metabolism in, 168
  chest shape in, 540
  colloidal osmotic pressure of plasma, 623
  new-born, prothrombin level in, 149
Infantile paralysis, 828
Inferior quadrigeminal body, 851
Inheritance, 1016
  of blood groups, 1017
  sex-linked, 1019
Inhibition, 839
  at motoneurones, 751
  competitive, 134
  of monosynaptic reflex, 750
Inhibitors, 73
  gastric, 245
  metabolic, 762
Initial heat, 761, 790
Injury, catabolic response to, 194
  current of, 743
Innervation of muscle cells, (fig.) *facing page* 773
  reciprocal, 845
Inositol, 13, 23, 136
Inspiratory capacity, 545, 547
  reserve volume, 548
Inspiration, 540
Insulation of ambient air, 210
  of clothes, 203
  of skin, 202
  thermal, 210
Insulin, 248, 301, 308, 311, 931
  actions, 934
  and hunger, 215
  and zinc, 933
  -antagonizing effect of adenohypophysial extracts, 300
  assay of, 933
  blood-sugar and, (fig.) 302, 303
  extraction of, 932
  hyperinsulinism, 935
  hypoglycæmia and, 311
  International Standard Preparation, 933
  liver fat content and, 291
  muscle glycogenesis, 303
  preparations, various and approximate times of action, 933, 934
  resistance, 294
  structure, 49, 932
  test of gastric secretion, 248
Intelligibility, 725
Intensity, minimum perceptible difference, 725
Intention tremor, 870, 872
Interalveolar cell-islets, 252, 301, 931
Intercostal muscles, 540
Intermediary metabolism, 283
Internal capsule, (fig.) 889
  ear, 714
  inhibition, 914
Interneurones, 736, 827
Internuncial neurones, 736, 753, 827
Interstitial cells, 951, 968
  -cell-stimulating hormone (ICSH), 969
  cells of testis, 951
  fluid, 629
    formation of, 636
Intestinal bacteria, 280
  digestion, 265
  enzymes, 265
  epithelium, (fig.) *facing page* 263, (fig.) *facing page* 272
  gas, 280
  glands, 264
  inhibition, 262
  juice, 264–267
    enzymes of, 265
  motility, 262
  tract, general structure of, 230
  villi, 263
Intestine, large, 276
  absorption, 268, 279
  bacteria of the, 280
  histamine in, 945
  hydroxytryptamine, 946
  in vitro, 260
  length of, 262
  size of, 262
  surface, 263
Intestine, small, absorption in the, 268
  digestive changes in, 265
  enzymes of, 264
  glands of, 264
  movements of the small, 260, (diagram) *facing page* 260
  secretion, control of, 267
Intracellular fluid, 629
Intracranial pressure, 647
Intra-ocular pressure, 492, 672
  tension, 671, 674
Intrapleural pressure, 544, 545
Intrathoracic pressure, 544
Intrinsic and extrinsic factors, 134, 237, 394
Inulin, 17, 594
  in kidney clearance, 604
Inversion, 14
Iodine, 106
  and endemic goitre, 927
    hyperthyroidism, 926
  daily intake, 928
  excretion of, 923
  in blood, 923
    thyroid gland, 920
  in water, 927
  radioactive and thyroid, 922
  table salt added to, 928
  uptake, and the thyrotrophic hormone, 923
  value, 21
Iodized table salt, 928
Iodo-acetic acid, 74, 313, 455

Ion exchange resins, 46
  exchangers, 38
  flux, 738
Iris, 670, 676
  functions of, 677
  innervation of, 676
Iron, 83, 84, 87, 97, 392, 404
  absorption, 392
  and hæmoglobin, 410
  content of common foodstuffs, (table) 97
  daily allowance, (table) 190
  deficiency anæmia, effect of iron on, (fig.) 393
  excretion, 99
  intake, 99
  pregnancy and, 983
  serum, 393
  transport, 99
    by placenta, 983
  pregnancy and, 983
Irradiation, 832
Ishihara's plates, 692
Islets of Langerhans, 252, 301, 308, 931, 1005
Isocitric dehydrogenase, 317
Isoelectric zone, 43, 50
Isoimmunization, 383
  Rh, 382
Isomerase, 812
Isometric lever, (fig.) 880
Isosmotic solutions, 37
Isosthenuria, 603
Isotonic solution, 37
  lever, (fig.) 779
  twitch, (fig.) *facing page* 779
Isotopes, 105, 284, 337
  chemical state of living matter and, 357
  in nitrogen metabolism, 337
  of carbon, (fig.) 285
  radioactive, 287
  stable, 285
Itching, 658

Jacksonian epilepsy, 892
Jam, 182
Jaundice, 383, 645, 658
  alkaline phosphatase in, 104
  and icteric index, 416
  hæmolytic, 416
  hepatic, 416
  obstructive, 257, 415
  toxic, 416
Jejunum, 260
Jenner's stain, 385
Joint impulses, 769
  position sense, 768
Jugular vein pressure, 487
  pulse, 439, 499
Junket, 241
Juxta-medullary glomeruli, 592

Kephalins, 23
Keratin, 48
Keratin-myosin-fibrinogen group, 48
Keratohyalin, 651
Keratomalacia, 118
Keto-acids, the fate of, 345
α-keto-glutaric acid, 317
β-keto-butyric acid, 290
Ketogenesis, 332
Ketogenic amino-acids, 333, 345
Ketolysis, 232
Ketonæmia, 332
Ketone bodies, 332
Ketonuria, 332
β-ketoreductase, 329
Ketosis, 8, 194, 332
17-ketosteroids, 27, 941, 952, 1004
  in cortical tumour, 953
  in urine, 953
Kidney, 97, 590
  aglomerular, 600
  ammonia formation and, 620
  and glutamine, 622
  and œdema, 620
  and pH, 620
  and posterior pituitary, 613
  and urinary pH, 642
  and water balance, 612
  appearance of, (fig.) *facing page* 592
  blood flow through the kidneys, total, 607
  blood vessels, 592, (fig.) *facing page* 593
  clearance tests, 604
    in man, 609
  denervated, pyrexial hyperæmia in, 610
  effect of hypertonic solutions, 615
  efficiency of, 600
  emotion and, 610, 619
  extracts of the neurohypophysis and, 613
  filtration pressure, 597
  glomerular function, 593
  hypertension, 610
  in acidosis and alkaline, 620
  inulin clearance, 604
  ischæmia, 610
  nephrectomy, 622
  œdema and, 620
  nerves, 592
  parathyroid hormone, 929
  pressure diuresis, 596
  regulation of hydrogen-ion concentration, 620
  renal function in the newborn, 622
  response to ingestion of water and salt, 612
  structure of, 590, (fig.) *facing page* 592
  tissue, estimation of active, 606
  tubles, 598
  vessels, (fig.) *facing page* 592
  volume and composition of urine, 601
  water balance, 612
  work of, 600
Kidneys and blood pressure, 610
Kinase, 76
Kilocalorie (kcal), definition, 152
Knee jerk, 839, 972
  in cerebellar injury, 872
  reflex inhibition of, (fig.) 840
Knoop, 283, 329
Korotkov sounds, 485
Krause's membrane, 771
Krogh's microtonometer, (fig.) *facing page* 585
Kupffer cells, 402
Kwashiorkor, 193

Labour, 989
  stages of, 991
Labyrinthectomy, 860

Labyrinthine reflexes, 855, (fig.) *facing page* 861, 862
  vertigo, 808
Labyrinths, 856
  connexions of vestibular fibres, (fig.) 858
  righting reflexes, 862
  unilateral removal, 860
Lacrimal glands, 670
Lactalbumin, 173
Lactase, 265
  in intestinal juice, 264
    pancreatic juice, 252
Lactation, 993
  and emotion, 994
  nutritional requirements, 996
  output of milk, (fig.) 994
Lacteals, 263, 639
Lactic acid, 122, 298
  anærobic contraction of muscle and, 777
  in blood, 799
  in cardiac metabolism, 410
  in muscles, 776
  muscle glycolysis, 311
  vasodilator substance, a, 516
Lactic dehydrogenase, 314
Lactogenic hormone, 993, 1003, 1006
Lactose, 13, 14, 174, 305
  in milk, 930
Lactobacillus casei, 125
  lactis, (Dorner), 133
Langerhans, islets of, 252, 931
Langhans' cell layer, 976
Lævulose, 13
Lag curve, 306
Large intestine, 276, (fig.) *facing page* 278
  absorption in, 279
  bacteria in, 280
  movements of, 276
  secretion, 279
  size of, 276
  structure of, 239
  transverse section, (fig.) *facing page* 280
Laryngeal nerve, 707
  tone, 706, 710
Laryngismus stridulus, 929
Larynx, 225, 705, 951
  nerve, 707
  structure, 706
Latent heat of vaporization, 203
Latent period, of synapse, 751
Lateral geniculate body, 899
  sulcus, 883
Law, 'all-or-nothing', 428
  Muller's, 648
  of specific irritability, 648
      nerve energy, 648
  of the heart, 446
  Poiseuille's, 500
  Starling's, 446
  Weber's, 648, 649
Laws of Grassmann, 690
  Pfluger, 755
Leads in clinical electrocardiography, 471
Lean tissue mass, 630
Lecithin, 22, 326
Lecithinase A, 22
Leech, 364
Leishman's stain, 385
Length units used, 1026
Lengthening reaction, 843
Lens, 673
Lentiform nucleus, 885
Leonardo da Vinci, 437
'Let-down', 994
Leucine, 283
Leucocytes, 396, 399
  basophil, 945
  differential count, 399
  effect of lungs on, 397
  enumeration, 399
  in blood at various ages, (fig.) 402
  shift to the left, 399
  survival, 397
Leucocytosis, 399
Leucopenia, 399
Leucotomy, 897
  prefrontal, 897
Leucovorin, 132
Lewisite, 74
Lewis's H-substance, 521
  triple response, 520
LH (luteinizing hormone), 968
Liebermann-Burchard test, 25
Life, maximum span, 1022
  of the red cell, 390
Ligature, Stannius, 427
Light and sexual activity, 970
  on skin, 522
  reflex, 677, 899
Liminal stimulus, 779
Linoleic acid, 194, 325
Linolenic acid, 194
Lipæmia, 274
Lipases, 20, 69, 265, 270
  in intestinal juice, 256
  in pancreatic juice, 253
Lipid, unsaponifiable fraction, 23
Lipids, 19–28
  in plasma, 324
  of tissue, 324
Lipoic acid, 124
Lipolysis, 270
Lipoproteins, 22, 54, 324, 371
Lipositols, 22
Lipotropic action, 327
  factors, 135, 327
Lipoxidase, 90
Liquor folliculi, 956
Litter weight, 991
Liver, 97
  amino-acid breakdown in, 340
  and androgens, 952
  choline in diet and, 327
  circulation through, 529
  cirrhosis of, 327
  destruction of œstrogens, 952
  detoxication mechanisms, 419
  fat in, 326
  function tests, 256, 296, 418
  functions, summary of, 418
  glycogen, 293
    disease, 311
  lobule, structure of, (fig.) 530
  microscopic structure, 336, 418, 530
  necrosis and hepatic artery, 529
  night blindness, 119
  parenchymal cells, 255
  plasma proteins in diseases of the, 336
  poisoning, 326
  secretions pressure, 258

Liver, temperature, 200
L-noradrenaline, 810
Local hormones, 945
Localization, 656
  in reflexes, 846
  of pain, 806
  of sound, 722
  of touch, 656
Locke's solution, 431
Long sight, 679
Loudness levels, 731
Lovén reflex, 514
Luciferase, 90
Lumbar puncture, 644, 645
Luminance, range of, 688
Lung, elasticity, 545
  dead space, 549
  lobules, 538, 539
  specific gravity, 551
Lungs, air passages of, 537
  and platelets, 400
  blood in, measurement of, 528
  ciliary activity, 538
  circulation through, 527
  effect on leucocytes, 397
  expansion of, 542
  giant cells of, 400
  in intrauterine life, 551
  inflation of, and cardiac acceleration, 497
  pleura, 539
  respiratory parts of, 538
  structure of, 537
  subdivisions of volume, (fig.) 547
  total capacity, 549
Luteal phase, 957
Lutein cells, 956, 968
Luteinizing hormone (LH), 968
Lymph, 638
  follicle, 260
  from the thoracic duct, 398
  nodes, 263, 638
    subcapsular lymph space, 639
Lymphatics, 263, 638
Lymphoblast, 398
Lymphocyte, survival, 398
Lymphocytes, 396, 398
Lysergic acid diethyl amide, 946
Lysine, 42, 185, 337, 351
Lyso-lecithin, 22

Macrophages, 398
Macula lutea, 682
  of utricle, 856
Macular vision, 898
Magnesium, 77, 106
  sulphate and bile, 260
Main d'accoucheur, (fig.) *facing page* 929, 929
Make 'shock', 759
Male sex hormones, 27, 951
  organs, (fig.) 949
Malic dehydrogenase, 318
Malleus, 711
Malpighi, 423
Malphigian bodies, 531
Malpighian corpuscle, 590
Maltase, 73
  in intestinal juice, 266
  in pancreatic juice, 252
Maltose, 13, 14
Mamillary bodies, 971
Mammary gland, 992
  and œstrogens, 993
  and progesterone, 993
    and suckling, 993
  blood flow in, 996
  metabolism, 996
Mammogen, 993
Manganese, 77, 107
Mannose, 12
Manometer (Wigger's), 435
  mercury, 484
Marchi's method, 763, 824
Marey's law, 464
Margarine, 21, 119, 177
  and vitamin A in, 119
Masking, 728
  binaural, 729
  monaural, 728
Mass reflexes, 830
  spectrometer, 286
  units used, 1026
Mast cells, 363, 397, 945
Mastication, 218
Maternal nutrition, 1021
Mating, behaviour, 963
Maximum expiratory effort, 550
  reabsorptive capacity, 599
Maximum breathing capacity, 550
  expiratory effort, 550
  reabsorptive capacity, 599
Mayer waves, 508
Meal, rate of passage, 262
Mean arterial pressure, 500
  blood pressure, 491
  corpuscular hæmoglobin, concentration, 392, 407
  corpuscular volume, 387
Meat, 177
Meconium, 280
Medial forebrain bundle, 876
  geniculate body, 722
Median plane, 819
Medulla, adrenal, 941
  oblongata, 847, (fig.) *facing page* 848
Medullary sheath, 733
Megacolon, 278
Megakaryocytes, 389, 400
Megaloblasts, 389, 390, 394
Meiosis, 1013
Meissner's plexus, 230
Melanin, 81, 349, 940
Melanocytes, 651
Melanophore-expanding hormone, 1007
Melanophores, 1007
Membrana granulosa, 956
Membrane potentials, 37, 736
  of living cells, 737
Menarche, 956, (fig.) 957
Mendelian laws, 382
  inheritance, 1017
Menopause, 960, 969, 972
Menstrual cycle, 957
  cycle, events in, (fig.) *facing page* 956
  daily excretion of œstrogens, (fig.) 966
    daily excretion of pregnanediol, 967
  flow, 958
  hormonal control of, 965
  temperature in, 966, 967, 974

Menstrual cycle, toxic factor in, 958
  withdrawal bleeding, 966
Menstruation and sebaceous activity, 652
  anovular, 959
2-mercaptoethylamine, 62
Mercapturic acids, 420
Mercury manometer, 484
Met, definition, 209
Metabolic acidosis, 643
  alkalosis, 643
  'constants,' (table) 154
  inhibitors, 74, 762
Metabolism, 2
  basal, 166
  carbohydrate, 288
  fat, 323
  inborn errors of, 284
  intermediary, 283
  interrelationships between carbohydrate, fat and protein, (fig.) 320
  nucleic acid, 355
  nucleoprotein, 355
  protein, 335
  sulphur, 354
  thyroid gland and, 923
  total, 168
Metabolite antagonists, 74, 762
Metaplasia, 118
Metaproteins, 55
Methæmalbumin, 409
Methæmoglobin, 405, 408, 586
  absorption spectra, (fig.) *facing page* 406
Metathalamus, 879
Methionine, 185, 347, 352, 354, 965
2-methyl-1: 4-naphthoquinone, 148
Methyl testosterone, 952
Methylene blue, 85
Micro-electrode, 735
Microsomes, 67, 78
Micropipette, 490
Microtonometer, (fig.) *facing page* 585
Micturition, 624
Mid brain, 847, 851
Middle ear, 711
Milieu intérieure, 3
Milk, 97, 173
  amount produced, 994
  carbohydrate in, 174
  cheese and, 176
  chemical composition of, 173
  clotting of, 174
  comparison of human with cow's, (table) 173
    different species, 175
  condensed, 176
  dried, 176
  -ejection response, 995
  iron content of, 97
  letting down, 994
  minerals in, 174
  pasteurization of, 175
  pigeon's, 1006
  precursors, 996
  production, 993
  proteins in, 173
  secretion of, 993
  sterilized, 175
  vitamins, 174
Milliequivalent, 29
Millimolar solution, 29
Milling, 179
Milliosmole, 29
Millon's reaction in protein, 55
Mineralocorticoids, 938
Minerals, 96
  absorption of, 269
Miner's nystagmus, 701
Minimum requirement, 114
Miosis, 677
Mitochondria, 67, 78, 316, 418
Mitosis, 1013
Mitral stenosis, 539
Modalities of cutaneous sensation, 659
Molar solution, 29
Molecular size of proteins, 45
  weights of proteins, 45
Molybdenum, 87, 107
Monaural masking, 728
Monochromats, 691
Monocytes, 396, 397
Monophasic action potential, 743
Monosaccharides, 8–13
Monosynaptic pathway, 826
  reflexes, 749, 839
Morphine, 678
Motion sickness, 861
Motoneurones, 733, 832
  of spinal cord, 750
Motor areas of cerebral cortex, 890
Motor aphasia, 896
  areas of cerebral cortex, 890
  cell of anterior horn, 733, (fig.) *facing page* 736, (fig.) *facing page* 745
  cortex, 888, 893
    effect of stimulation, 891
  decussation, 848
  end-plate, 771
  homunculus, 891
  neurone, after discharge, 833
  paralysis, 829
  unit, 782
Modiolus, 714
Mouth, 218
  absorption in, 268
  cavity, 706
  temperature, 200
M.R.C. hæmoglobinometer, 406
Mucin, 18, 222, 224, 237
Mucoitin sulphate, 18
Mucopolysaccharides, 17
Mucoproteins, 18
Mucous cells, 224
Mucus, 241
Müllerian ducts, 948
Muller's law, 648
Muscle, 771
  action potentials, (fig.) *near page* 785
  aerobic recovery, 778
  all or nothing law, (fig.) *facing page* 784
  antigravity, 841
  ATP and ADP in, 774
  capillaries of, 518
  carbohydrate metabolism in, 776
    utilization, 776
  cardiac, 425
  chemistry of, 772
  contraction and ATP, 775
  contraction of, characteristics, 779
    experiments on frog's muscle-nerve preparation, 779
    grading of, 783

Muscle, contraction of, lactic acid, 776
    mechanism of, 775
    processes involved, three, 777
    threshold or liminal stimulus and, 779
    voluntary, 783
  creatinine excretion and, 353
  crushing, 774
  effect of temperature, 782
  electromyography, 792
  electronmicrograph, (fig.) *facing page* 772
  end-plate, 773, 788
  excitability of, 756
  extractives, 774
  exercise, 793
    cardiac output and oxygen consumption, approximate figures, (figs.) 800
    chemical changes in, 796
    circulatory and respiratory adjustments, 793
    oxygen debt, 798
    respiratory quotient during and after, (figs.) 795
    second wind, 800
    submarine analogy, 796
    fatigue in, 776, 786
      in severe exercise, 787
  glycogen in, 298, 774
  heart, 802
  heat and tension, (fig.) 781
    production in, 790
  injury and, 774
  innervation of, (fig.) *facing page* 773
  -joint sense, 827
  motor unit, 782
  myoneural junction, 787, 788
  oxygen requirement, 584
  pain, 772
  plasma, 773
  nerves to, 771
  proteins of, 773
  pump, 436, 523
  red and white, 771
  refractory period, 782
  skeletal, 771
  smooth, 801
  spindle, 768, 771, 772, 841
  strength, 1024
  striated, 771
  structure of skeletal, 771
  temperature in exercise, 797
  tension, 835
  tone, 784
  tremor in hyperthyroidism, 926
  twitch, 779
  vasodilatation, 516
  visceral, 801
  viscosity, 784, 786
  work, 785
    flywheel for measuring, 785
Mustard gas and mutation, 1020
Mutarotation, 9
Mutation, 1020
Myasthenia gravis, 944
Myelin sheath, 733
Myelination, 889
Myelocytes, 397
Myeloid series of white blood cells, 396
Myenteric nerve plexus, 230
  reflex, 261
Myoepithelial cells, 996
Myofibrils, 771
Myoglobin, 774
Myohæmoglobin, 774
Myometrium, 954
Myoneural junction, 788
Myopia, 679, 680
Myosin, 773
Myotatic reflex, 749, 839, 843
Myxœdemia, 395, 920, 924, (fig.) *facing page* 925, 1002
  anæmia in, 395
  basal metabolic rate in, 924
  blood cholesterol in, 333
  extract of sheep's thyroid in, 924

Nasal air stream, 536, 662
  glands, 662
  mucosa, 536
Native proteins, 52
Near point, (fig.) 676
  reflex, 678
Neck reflexes, 855
Negative after image, 691
  conditioned reflex, 913
  successive induction, 845
  supporting reaction, 862
Neocerebellum, 865, 873
Neopallium, 883
Neothalamus, 879
Nephrectomy, 622
Nephron, 590
Nephrosis, 324
Nerve, action potentials recorded from a cathode-ray oscilloscope, (fig.) 740, 744
  all or none law, 739
  adrenergic, 810
  cardio-aortic, 463
  conduction in, 735–744
  cholinergic, 812
  degeneration, 763, 822
  excitability in, 745, 755
  fatigue, 760
  fibre, rate of growth, 764
  fibres A, B and C, 744
    characteristics depending on diameter, 744
  gases, 74
  heat production and metabolism of, 761
  impulse, 739
    antidromic, 739
    electrical changes accompanying, 742
    humoral theory, 810
  injuries, 758
  ionic exchanges, (fig.) *facing page* 737
  metabolism, 761
  myelinated, 733
  non-myelinated, 735
  Pflüger's law and, 755
  plexus of submucosa, 230
  presacral, 277
  R. Q., 762
  refractory period, 744
  regeneration of, 763
  saltatory conduction, 740
  stimulation in man, 758
  threshold, 739
Nerves, 'buffer', 418
  from viscera, 807
  human, stimulation of, 759

Nerves, of colon and rectum, 277
  pelvic splanchnic, 277
  splanchnic, 261
  vasoconstrictor, 509
  vasodilator, 510
  vasomotor, 509
Nervi erigentes, 975
Nervous regulation of the heart, 457
  pathways, methods of determining, 822, 827
Nervous system, 817
  afferent impulses from the viscera, 805
  a homoiostatic mechanism, 815
  autonomic, 804
  metabolism of, 820
  nomenclature, 817
  parasympathetic outflow (thoracolumbar), 805
  scheme showing connexions of several parts of brain, (fig.) 818
  sympathetic, 805
Net efficiency, 165
Nettle-rash, 945
Neurilemmal sheath, 733
Neurohypophysis, 998, 999, 1009
Neuromuscular block, 787
  junction, 788
  transmission and curare, 749, 790, 930
Neurone, 733
Neurones, internuncial, 736, 753, 827
Neuropathy, 122
Neuropraxia, 763
Neurosis, experimental, 916
Neurotmesis, 763
Neutral fat, 326
Neutrophil polymorph, 397
Newborn, renal function in, 622
Nicotinamide, 61
Nicotinamide nucleotides, 61, 85, 127
Nicotine action on synapse, 509, 748, 804
Nicotinic acid, 61, 127, 130, 135, 284
  amide, 61, 127
  and coenzyme I and II, 127
  and pellagra, 128
  daily allowance, (table) 190
  deficiency symptoms, 128
  in foods, (table) 129
Night blindness, 118
Nissl's granules, 733, 765
Nitrogen balance, 185, 187, 336
  biological value, 189
  equilibrium, 185, 336
  excretion, 358, 603
    premortal rise, 195
  isotopes, 286
  metabolism, 335
  requirement, 192
Nociceptive response, 832
Nocifensor nerves, 659
Node, atrioventricular, 424, 425
  of Ranvier, 733, 740
  sinu-atrial, 424, 425, 454
Noise, 730, 731
  distribution throughout auditory sensation area, (fig.) 731
  effects of, 730
  ladder, 731
Nominal aphasia, 904
Non-myelinated fibres, 735
Non-polarizable electrodes, 754
Non-protein, respiratory quotient, 156
Noradrenaline, 221, 304, 306, 516, 753, 941
  and adrenaline effects of, (table) 811
  content of organs, 812
  estimation of, 942
  in blood, 944
  in urine, 943
  on heart, 462
Normal diet, 184
  solution, 29
Normoblast, 390
Normocytes, 388
Nose, 536
  and sense of smell, 661
  in respiration, 536
  olfactory area, 661
  route taken by stream of air, 536, 662
Nucleases, 253, 265
  in intestinal juice, 264
  in pancreatic juice, 252
Nucleic acid, 54, 57, 63, 355
    digestion of, 267
  and nucleoprotein, metabolism, 355
Nucleolus, 64, 66
Nucleoprotein metabolism, 355
Nucleoproteins, 54, 57, 355
Nucleoside, 59
Nucleotidase in intestinal juice, 264
Nucleotide, coenzymes, 60
Nucleotides, 60
Nucleus, 66
  cuneatus, 880
  dentatus, 867
  emboliformis, 867
  fastigii, 867
  globosus, 867
  gracilis, 880
  supraoptic, 876
Neutral sulphur, 420
Nutrient enemata, 279
Nutrition, 173, 197
Nutritional anæmia, 99
  factors in hæmopoiesis, 395
  muscular dystrophy, 147
Nuts, 182
Nystagmus, 700, 701, 870, 875
  optokinetic, 861
  vestibular, 861

Oatmeal, 181
Obesity, 215, 879
Obligatory reabsorption, 598
Obstructive jaundice, 257, 415
Occipital cortex, 898
Occlusion, 748, 835
Occupation and calorie requirements, 170
Ocular manifestations of vitamin deficiency, 120
Odour, 663
Œdema, 193, 373
  and kidney, 620
  famine and, 373
  formation and heart failure and, 620
  in beri-beri, 124
  kidney and, 620
Œsophageal speech, 708
Œsophago-salivary reflex, 226
Œsophagus, 221
  mucosa, 229
  physiology, 226

Œsophagus, salivation, 226
  sphincter, 228
  structure of, 225
Œstradiol, 27, 961, 968
Œstriol, 961
Œstrogen and menstrual cycle, 966
  climacteric and, 972
  destruction of, 965
  excretion, (table) 953
  excretion in menstrual cycle, 966
  excretion in pregnancy, (fig.) 977, 978
  from testis, 953
  inactivation by liver, 373, 965
  in adrenal cortex, 938
  mammary gland and, 993
  synthetic, 964
  uterine muscle activity, 963
Œstrogens and mammary gland, 993
  and pituitary inhibition, 970
Œstrone, 27, 961
Œstrous cycle, 960, 970
Œstrus, 960
Old age, changes in, 1023
Oleic acid, 19
Olfaction, 661
  adaptation in, 663
  and pigmentation, 662
Olfactory bulb, 663
  epithelium, (fig.) 662
  tract, 663
Oligæmia, 524
Oncotic pressure, 38
Oöcyte, 1014
Oöphorectomy, 965
  and pregnancy, 977
Ophthalmoscope, 683
Ophthalmoscopic view of eye, (fig.) *facing page* 680
Optic disc, 685
Optic chiasma, (fig.) *facing page* 896, 898
  radiations, 898
Optic pathways, 898
Optics of the eye, 672
Optimal requirement, 114
Optimum pH, 71
  temperature, 71
Optokinetic nystagmus, 861
Ora serrata, 669
Oral cavity, 706
  temperature (histogram), 208
Orbicularis oculi, 670
Organ of Corti, 714
Orgasm, 974
Origin of blood cells, 388
Ornithine, 342, 351
Ornithine-arginine cycle, 342
Orthopnœa, 549
Osazone, 11
Oscillograms of vowels, 709
Osmic acid, 823
Osmoreceptors, 617
Osmotic diuresis, 600
Osmotic pressure, 34–36
  and thirst, 216
  and urine output, 616
  of plasma, 372
Ossein, 104
Ossification, 105
Osteitis fibrosa, 101
Osteoblasts, 144
Osteomalacia, 145
Otic ganglion, 219
Otolith organ, 858
Otoliths, 856
Otosclerosis, 712, 727
Output of the heart, 445, 450
Ovarian hormones, 961
Ovary, (fig.) *facing page* 956, 957
  and pregnancy, 977
  follicles (Graafian), 968
  hormones, 961
Ovulation, 956, 970
  and basal temperature, 974
  and fertility, 972
  FSH and LH in, 969
  time of, 973
Ovum, 1014
  fertilized, 975
  viability, 972
Oxalate and clotting, 362
Oxaloacetic acid, 317, 318
Oxidases, 69, 80
  reaction, 897
β-Oxidation, 329
Oxidation mechanism of, 80
Oxidation-reduction potentials, 90
Oxidative deamination, 341
  phosphorylation, 318
α-Oxoglutarate, 317
Oxoisomerase, 311
17-Oxosteroids, 27, 941, 952, 1004
Oxygen, bradycardia, 560
  calorific value of, 154
  capacity of blood, 573
  content of blood, 573
  debt, 165, 778, 798
  dissociation curve, 573
  effect of increased, 558, 559
  in blood, 572
  lack, 557, 585
    at high altitudes, 557, 585
  requirement, 544
    at various ages, (table) 544
  saturation, 573
    of fœtal blood, 985
  tension, arterial and venous, (table) 572
  therapy, 588
  transfer in placenta, 982
    BLB inhalation apparatus, (fig.) 588
  transport of, 573
Oxycalorimeter, 153
Oxyhæmoglobin, 404
  absorption spectra, (fig.) *facing page* 406
  dissociation curve of, 575
Oxyntic cells, 231, 238, 239, 246
Oxytocin, 960, 963, 975, (fig.) *facing page* 1009, 1009

Pacemaker, 428, 454, 458
Packed cell volume, 387
Pacinian corpuscles, 806
Pain, angina pectoris and, 808
  diaphragmatic, 808
  fibres in sympathetic nerves, 805
  gall-bladder, 808
  heart disease, 808
  minimum stimulus, (table) 657
  muscle, 772

Pain, nocifensor system, (table) 659
  projection of, 806
  sense, 657
  sense and asphyxia, 658
  testis, 808
  visceral, 805
Palæocerebellum, 865
Palæostriatum, 884
Paleothalamus, 879
Palate, 225
Pallium, 883
Palmar erythema, 965
Palmitic acid, 19
Pancreas, 251, (fig.) *facing page* 254
  diabetes mellitus and, 307
  islets of, 931, 1005
  juice, 252
  microscopic section, (fig.) *facing page* 254
Pancreatectomy, 302, 307, 932, 934
Pancreozymin, 254, 255
Paneth cell, *facing page* 260
Pantothenic acid, 62, 130, 135
Pantoyltaurine, 135
Paper chromatography, 46
  electrophoresis, 52
Papillœdema, 647
Parahæmatin, 411
Parallax, 702
Paralutein cells, 956
Paralysis agitans, 886
  flaccid, 894
  organic motor, (table) 829
  spastic, 894
Paralytic secretion, in salivary gland of dog, 221
Paraplegia, 627, 815, 830
  urinary bladder in, 815
Parasympathetic outflow, 805
Parathormone, 929
Parathyroid, activity, unit of, 929
  extract and rise of serum calcium concentration, 929
  glands, 101, 928
  histology, (fig.) *facing page* 928
  PTH (parathyroid hormone), 929
  removal of, and tetany, 929
  tetany, (fig.) *facing page* 929
  tumour of, 931
  tumour and bones, (fig.) *near page* 932
Parathyroidectomy, carpopedal spasm, (fig.) *near page* 929
  tetany in, 929
Parietal cells, 236, 238
  lobes, 902
  pleura, 539
Paraventricular nucleus, 1000
Parotid gland, 218
Paroxysmal tachycardia, 464
Partial pressure, 571
Partition hypothesis of fat absorption, 271
Parturition, 989
  and uterine distension, 991
  hormonal basis of, 990
Past-pointing, 861
Pasteur effect, 316
Pasteurization, 175
Pavlov pouch, 235, 238, 243, 244
Pedal spasm, (fig.) *facing page* 929
Peduncles, cerebellar, 865
Pellagra, 127
Pelvic splanchnic nerves, 277
Pentoses, 13
Pentosuria, 284, 305
Pepsin, 237, 238, 241, 246
Pepsinogen, 238, 241
Peptic cells, 238
Peptic digestion, products of, 241
Peptides, 44
Peptones, 55
Perception, 648
Pericardium, 426
Perhydrocyclopenteno-phenanthrene, 24
Perceptive deafness, 730
Perilymph, 714, 856
Perimeter, 687
Perimetry, 898
Perineurium, 737
Peripheral circulatory failure, 524
  resistance, 499
  resistance and adrenaline, 517
Peristalsis, 261, 232, 276
  gastric, 234
Peritoneum, parietal, 808
Pernicious anæmia, 133, (fig.) *facing page* 386, 394
  red cells in, (fig.) *facing page* 388
Perosis, 327
Peroxidase, 83, 98
Perspective, 702
Petechiæ, 140
Peyer's patches, 263
Pflüger's laws, 755
pH (hydrogen-ion concentration), 31
  of blood, 641
  of gastric juice, 246
  of the urine, 642
  urine, 621
  notation, 31
Phaeochromocytoma, 944
Phagocytes, 402
Pharyngotympanic tube, 712
Phenol and liver function, 419
  oxidase, 73, 77, 81
Phenotype, 1017, 1018
Phenylalanine, 185, 348
Phenylketonuria, 349
Phenylthiourea, 1017
Phlebogram, 498
Phlorrhizin, 284, 307, 605
  and glycosuria, 284
Phon, 730
Phonation, 706
Phonocardiogram, 438, 439, (fig.) 474
Phosphagen, 775
Phosphatase, 104, 106, 175, 193, 265, 294
  prostatic, 950
  test, 175
Phosphate, 102
  and vitamin D, 144
  bond, energy-rich, 60, 313, 318, 776
  serum, 929
Phosphatides, 21
Phosphatidyl serine, 23
Phosphocreatine, 455, 755
Phosphodihydroxyacetone, 312
6-phosphofructose, 312
Phosphoglucomutase, 294
1-phosphoglucose, 11, 295
6-phosphoglucose, 11, 295, 311
3-phosphoglyceraldehyde, 312

Phosphoglyceraldehyde dehydrogenase, 90, 312
Phosphoglyceromutase, 313
Phosphohexokinase, 312
Phosphohexose isomerase 311
Phosphokinase, 313
Phospholipids, 21, 326
Phosphomonoesterase, 104
Phosphoproteins, 54, 102
Phosphopyruvic acid, 313
Phosphorolysis, 294
Phosphorus, 102
  absorption, 103
  bones and, 104
  compounds in blood, (table) 103
  content of foods, 100
  excretion, 103
  in pregnancy, 983
  requirement, 103
  teeth and, 104
Phosphorylase, 294
Photoelectric colorimeters, 406
Photopic vision, 685, 693
Physiology, definition, 1
Physostigmine, 74, 813
Phytic acid, 13, 102, 136
  cereals and, 178
  content of foods, (table) 102
Pia mater, 644
Pial vessels, 534
Pigeon's milk, 1006
Pigmentation and olfaction, 662
Pigments of blood, 404–416
Pilocarpine and pancreas, 253
  salivary secretion, 221
Pineal body, 944
Pinhole camera, 681
Piqúre diabetes, 307
Pitch, 725
Pituicytes, 999
Pituitary, anterior lobe, 1002
  acromegaly, 1007
  adenohypophysis, 1002
  adrenocorticotrophic hormone ACTH, 1003
  antidiuretic hormone, 613, 1010
  blood supply, 1000
  body, 998
  body, terminology, 998
  carbohydrate metabolism and, 300
  dwarfism, 1008
  gland of dog, (fig.) *facing page* 1000
  gonadotrophic hormones and, 1005
  nerves of, 999
  neurohypophysis, 999
  growth hormone, 1005
  lactogenic hormone, 1006
  thyrotrophic hormone, TSH, 1002
Placenta, 981
  amino-acid transfer, 983
  and hormones, 977
  circulation in, (fig.) *facing page* 984
  formation of, 976
  iron transfer, 983
  oxygen transfer, 982
  sugar transfer, 982
  vitamin transfer, 983
  weight of, 981
Placido's disc, 673
Placing reactions, 863
Plantar reflex, 830
Plasma, 103, 360
  albumins, 369, 371
  and urine composition compared, 593
  calcium, 100
  clearance, for inulin, 604
  globulins, 369, 371
  hæmophilia and, 366
  iodine, 924
  lipids, 324
  osmotic pressure, 372
  oxalated, 362
  salted, 363
  separated, 577
  skimming, 377
  true, 577
  volume, 377
    in pregnancy, 980
Plasma proteins, 193, 336, 338, 361, 369
    concentration, 370
    electrophoresis apparatus (Tiselius), 51, (paper) 52, 56
    formation of, 373
    functions of, 373
    hypoproteinæmia, 371
    molecules of, 372
    osmotic pressure, 372
Plasmapheresis, 193, 373
Plasmin, 367
Platelets, 368, 400, 946
  and spleen, 400
  disintegration, 365
  enumeration, 400
  in coagulation, 366
Plethysmograph, 502, 503
Pleura, 539
Plexus, myenteric, 230
  of submucosa, 230
Pneumotaxic centre, 554
Pneumothorax, 546
Poikilothermic animals, 198
Poiseuille's Law, 500
Polar body, 1015
Polarized membrane, 737
Polycythæmia, 386, 588
Polyhexozamines, 17
Polymorphonuclear cells, 397
Polynucleotides, 63
Polyneuritis, 122
Polypeptide, 45, 55
Polysaccharide, 8, 15–17
Polysynaptic reflexes, 749
Polyuronides, 17
Pons, 847
  hæmorrhage into, 206
Porphin, 411
Porphobilinogen, 412
Porphyria, 413
  congenital, 368, 413
Porphyrin, 411
Portal hypertension, 530
  system, 529
  venous pressure, 529
Positive after-image, 691
  supporting reaction, 846
Postcalcarine sulcus, 898
Postcentral gyrus, 902
Posterior root fibres, 826
Postganglionic fibres, 804
Postural reflexes, 847, 862
  syncope, 523

Posture and decerebrate rigidity, 852
Potassium, 107
  and electrocardiogram, 475
  transfer, 739
  in blood, 107
  in cells, 737
  space, 633
Potato, 141
  ascorbic acid in, 141
Potential across membrane of cells, 737
  demarcation, 743
  resting, 737
  reversal, 737
Pouch, Pavlov, 235, 238
Precentral gyrus, 884
Precipitin test, 55
Prefrontal areas, 897
  leucotomy, 897
Preganglionic fibres, 804
Pregnancy, 969, 974
  and gonadotrophin excretion, 980
  and œstrogen excretion, 978
  and placenta, 977
  and pregnanediol excretion, 979
  and sebaceous activity, 652
  and vitamin E, 147
  calcium and phosphorus requirements in, 983
  cessation of menstruation and, 975
  conception, 974
  diagnosis tests, 978
  endometrium, (fig.) *facing page* 972
  hormonal control of, 977
  iron in, 983
  lactation and, calorie requirements in, 996
  plasma volume in, 980
  proteinuria and, 595
  pseudo-, in rabbit, 973
  test, xenopus, (fig.) *facing page* 977
  urinary excretion of hormones in, 977
  water retention, 980
  weight gain, 981
Pregnane, 26
Pregnanediol, 28, 964
  excreted during pregnancy, 977
  excretion, 967, 979
  in urine, 953
Pregnant mare serum, 969
Premotor area, 895
Presbycusis, 727, 732
Presacral nerve, 277
Presbyopia, 676
Pressure, arterial, (fig.) *facing page* 457, 483
  atrial, 433, 451, (fig.) *facing page* 488
  capillary, 490
  colloid osmotic, 37
  diuresis, 596
  effective filtration, 597
  in aorta, 433
  in brachial artery, (fig.) *facing page* 528
  in brachial and posterior tibial artery, (fig.) 487
  in salivary gland, 223
  in vessels, 483
  intrapleural, 544, 545
  mean blood, 491
  osmotic, 38
  portal venous, 529
  pulmonary artery, 527
  pulmonary capillary, 527
Pressure, pulse, 492
  venous, 487
  ventricular, 433 (fig.) *facing page* 457 and 528
Pretectal area, 899
Price-Jones' curve, 385, (fig.) *facing page* 386
Prickle cells, 651
Primary lobule of lung, 538
Progestational proliferation, (fig.) *facing page* 964
Progestens, 27
Progesterone, 27, 964, 966, 967, 974, 977
  and mammary gland, 993
  on uterine movements, (fig.) *facing page* 965
Projection, 656
  fibres, 888
  of chemical senses, 661
  of pain, 806
Prolactin, LTH, 993, 1003, 1006
Proliferation, phase of, 958
Pronation-supination, (fig.) 874
Prone-pressure method of artificial respiration, 567
Pro-œstrus, 961
Proptosis, 677
Proprioceptors, 768
Prostate gland, 104, 950
Prostatic secretion, 950
Prostatic phosphatase, 950
Prosthetic group, 76
Protamine, 53, 65, 933
  and clotting, 363
  zinc insulin, 933
Protanopes, 692
Protanopia, 693
Protein, amino-acids of, 41
  anabolism and androgens, 954
  and hard work, 192
  -bound iodine, 923
  calorie value of 154,
  climate and, 202
  correction in metabolism, 156
  daily allowance, (table) 190
  deficiency, 193
  denaturation, 52
  hydrolysates, intravenous administration of, 335
  in urine, 595
  intake and nitrogen equilibrium, 187
  keratin-myosin-fibrinogen, 48
  metabolism of, 335
  molecular size of, 45
  requirements of, 184, 192
  storage, 335
  sparing action of carbohydrate and fat, 189
  synthesis, 335
Proteinases, 69
Proteins, 40–56
  absorption of, 269
  biological reactions of, 55
    value of, 189, 1021
  blood, 374
  classification of, 53
  collagen group of, 49
  colloidal properties of, 50
  conjugated, 54
  corpuscular, 48, 49
  denaturation of, 52
  digestion of, 265
  digestive derivatives of, 55

Proteins, fibrous, 48
  general reactions of, 54
  globular, 48, 49
  half life of, 339
  isoelectric point, 50
  molecular weights of, (table) 46
  of plasma, 369, 374
  reactions of, 54
Proteinuria, 595
Proteoses, 51
Prothrombin, 362, 365, 421
  consumption test, 366
  time, 149, 366
    and vitamin K, 148
Protoporphyrin, 412
Provitamins, 116
Pruritus, 658
Pseudolarynx, 708
Pseudopodia, 400
Pseudopregnancy, 961
Psychic secretion, 242
Ptosis, 677
Pteroyl-glutamic acid (PGA), 132
Ptyalin, 222, 246
Puberty, 969
  delayed, 954
  hormonal basis of, 971
  in boy, 951
  in girl, 956, 993
  precocious, 970, 971
Pudendal nerve, 277
Puerperium, 992, 993
Pulmonary artery pressure, 527
  circulation, 527
  vascular resistance, 528
  ventilation, 547
Pulse, arterial, 493
  dicrotic, 495
  jugular, 439, 499
  pressure, 486, 492
  rate, 467, 495
  venous, 497
  wave, shape of, 494
Pulse-wave, initiation of, (fig.) 494
  velocity, 493
Pupil of eye, 677–678
Pupillary reflexes, 899
Purine bases, 57, 355
  metabolism of, 355
Purkinje cells, 453, 866
  tissue, 453
Purkinje's experiment, 685
Purkinje-Sanson images, 675, (fig.) *facing page* 681
Putamen, 884
Putrescine, 351
Pyloric antrum, 231
  glands, 239
  sphincter, 231, 234
Pylorus, 231
Pyramidal tracts, section of, 894
Pyramids, 848, 849, 854, 890
Pyran, 10
Pyranose, 10
Pyrexia, 209
Pyridine, 407
  hæmochromogen, 407
  nucleotides, 61, 85
Pyrexia, artificial, 610
Pyridoxal, 129
Pyridoxal, phosphate, 129, 341, 350
Pyridoxamine, 129
Pyridoxine, 129
Pyrimidine bases, 57
  metabolism of, 355
Pyrithiamine, 135
Pyrrol rings, 410
Pyruvic acid, 122, 313, 316
  blood pyruvate following exercise, (fig.) 777
  disposal of, 313

Quantum, 698
Queckenstedt's sign, 645
Quinidine, 479

Radiation, 202
Radioactive tracers, 738
Ranvier, node of, 733
Rate of the heart, 467
Reabsorptive capacity, 599
Reaction of degeneration, R.D., 759, 829
Reactions of denervation, 759, 829
Reactive hyperæmia, 515, (fig.) *facing page* 520
Reading, 700
Rebound of motoneurones, 839
Recessive character, 1016
Reciprocal innervation, 845
  reflex, (fig.) 843
Recovery heat, 761
Recruitment, 838
  of neurones, 836
Rectum, 277
  nerves, 277
  temperature, 199, 212
Recurrent nerve, 707
Red blood cells, 385
  cell diameter, (fig.) *facing page* 385
    life of, 390
    in pernicious anæmia, (fig.) *facing page* 390
    sedimentation, 389
  cell, mean volume, 387
Red cells and altitude, 386
  in blood at various ages, (fig.) 402
  number of, 385
  origin of, 386
  volume, 376
Red nucleus, 851, 853, 868
Redox potential, 91
Reduced latent period, 751
  reflex time, 832, 839
Reduction, 80
  division, 1014
Reflex arc, 6
  accommodation, 674, 901
  action, definition, 817
  axon, 512
  conditioned, 911
  enterogastric, 235
  extension, 837
  flexor, 750, 830
  gastrocolic, 279
  grasp, 896
  Hering-Breuer, 562
  light, 899
  localization, 846
  Lovén, 514
  monosynaptic, 749
  myenteric, 261

Reflex arc, œsophago-salivary, 226
  plantar, 830, 894
  polysynaptic, 749
  pupil, 899
  scratch, 845
  unconditioned, 912
Reflexes, after-discharge, 833
  fatigue, 836
  from great veins and right atrium, 466
  heart, 462
  labyrinthine, 855, 862
  neck, 855
  postural, 847
  righting, 862
  spatial summation, 833
  spinal, 830
  stretch, 839
  vasoconstrictor, 512
  vasodilator, 512
  visual, 898
  withdrawal, 832
Refraction, 673
Refractive errors, 681
Refractory period, 428, 739, (fig.) *facing page* 744, 745, 782
  of heart muscle, 428
    nerve, 744
    skeletal muscle, 782
    synapse, 751
Regeneration of nerve fibres, 763
Reinforcement, 912
Rein's stromuhr for measuring blood flow in animals, 501
Remaining sensibility, 828
Renal clearance tests, 604, 609
  corpuscle, 590
  function in newborn, 622
  glycosuria, (fig.) 306
  hypertension, chemical mechanism of, 611
  threshold, 599
Renin, 612
Rennin, 173, 237, 241
Repair of nerve fibres, 764
Reproduction, 2, 948–997
Reproductive activity and seasonal variation in animals, 970
Residual volume, 547
Resistance, peripheral, 499
  pulmonary vascular, 528
Resonance theory, 715
Respiration, 536
  and exercise, 565
  artificial, 566
  central control of, 553
  chemical regulation of, 555
  Cheyne-Stokes, 559
  deglutition and, 226
  effect of O_2 lack, 557
  fœtal and neonatal, 551
  movements of, 540
  nervous regulation of, 561
  newborn, in, 551
  periodic, 559
Respiratory acidosis, 644
  alkalosis, 644
  arrhythmia, 497
  bronchioles, 538
  centre, (fig.) *facing page* 552, (fig.) *facing page* 553
    stimulation of cells, 560
Respiratory acidosis, efficiency tests, 550
  flow rate, 550
  function of blood, 571
  minute volume, 547
  movements, recording of, 542
  pump, 436
  quotient and calories per litre of oxygen (nomogram), 155
    definition of, 154
  quotient during and after muscular exercise, 795
  quotient of nerve, 762
  rate, 544
    at birth, 551
    at different ages, (table) 544
  tests, 550
Rest pauses, 172
Resting potential, 736, 737
  of motoneurones, 750
Resuscitation, 586
Rete testis, 949
Retention incontinence, 627
Reticular formation, 850, 852
Reticulocytes, 385, 390, 393
Reticulo-endothelial cells, 398
  system, 389, 402, 414
  tissue, 371
Retina, 118, 681–685
  corresponding points, 675
  distance between retinal vessels and rods and cones, 684
  electrical activity of, 694
  grouping of neurones into functional systems, (fig.) 683
  light-sensitive layer of eye, 684
  number of rods and cones, 684
  Purkinje's experiment, 684
  structure of, 681
  visual acuity, 685
Retinal directional effect, 691
  interaction, 699
Retinene, 119
Retinitis pigmentosa, 1019
Retrolental fibroplasia, 560
Reverberation, 753
Reversal of potential, 737
Reversion spectroscope, 408
Reynold's number, 500
Rh factor, 382
  -negative, 382
  -positive, 382
Rheobase, 755
Rhesus factor, 382
Rhinencephalon, 819, 883
Rhodopsin, 119
Rib movements, 541
Riboflavin, 125, 965
  -adenine-dinucleotide, 126
  daily allowance, (table) 190
  deficiency, (fig.) *facing page* 129
  in foods, (table) 127
  phosphate, 86, 126
Ribonuclease, 63, 66
Ribonucleic acid (RNA), 57, 418
Ribonucleoprotein, 64
Ribose, 13, 57
Rickets, 142, 144
  external appearances, (fig.) *facing page* 145
  radiological appearance, (fig.) *facing page* 144

Right atrial reflex, 466
Righting reflexes, 862
  eye, 863
Rigidity, decerebrate, 852
  decerebrate in man, 863
Rigor mortis, 778
Rima glottidis, 705
Ringer's solution, 431
Rinne's test, 730
Riva-Rocci method, human blood pressure, 484
RNA, Ribonucleic acid, 57, 63, 64
Rocking method of artificial respiration, 569
Rod monochromats, 692
Rods, number of, 684
  of retina, 682
Romanowsky stains, 385
Rotational movement, 858
Roth-Benedict apparatus, 159
Rouget cells, 517
Roughage, 182, 282
R.Q., respiratory quotient, 155
Rubner's values, 152
Rugæ, 231
Ryle's tube, 237

Saccadic movement, 700, 861
Saccule, 856
Saggital plane, 819
Sahli hæmoglobinometer, 406
Saliva, composition of, 222
  formation from blood, (fig.) 223
  functions of, 222
  mechanisms of secretion, 223
  psychic secretion, 221
Salivary acinus, 219
  amylase, 222
  calculus, 222
  digestion, 222
  glands, 218
    histology, 219, 224
    pressure in duct, (fig.) *facing page* 220, 223
    nerve supply of, 219
  secretion, histology of, 224
Salivation, 912
  and cellular changes, (fig.) 224
  in dogs and conditioned reflexes, 897
  nervous control, 221
Salt depletion, (fig.) 635
  intake and kidney response, 612
  sweating, 205
  taste, 667
Saltatory conduction, 737, 740
Saponification, 20
  value, 20
Sarcolemma, 771
Sarcoplasm, 771
Sarcostyles, 771
Satisfaction and hunger, 215
Saturation and vitamins, 114
  oxygen, 573
Saturation tests, 114
Sayers' method, 1004
Scala tympani, 713
  vestibuli, 713
Scanning speech, 875
Schafer's method of artificial respiration, (fig.) 567
Schizophrenia, 292
Schoenheimer, 337
Schwann cell, 733
Sclera, 670
Scleroproteins, 53
Scotopic vision, 685, 693
Scratch reflex, (fig.) 845
Scurvy, 109, 136, 140, 395
Scutellum, 179
Sea sickness, 861
Seasonal variation in reproductive activity, 970
Sebaceous secretion, 951
Sebum, 652
Second wind, 800, 801
Secondary lobule of lung, 539
Secretagogues, 243
Secretin, 254, (fig.) *facing page* 254, 255, 256
Secretion, histological basis of, 224
  in large intestine, 279
  of small intestine, 267
Secretors, 383
  non-, 383
Secretory phase, 958
Sedimentation rate, 389
Segmentation, 260
Semantic aphasia, 902
Semen, 950
Semi-circular canals, 856
    position of, (fig.) *facing page* 860
  ducts, 856
Semilunar valves, 426
Seminal fluid, 104, 950
  vesicle, 950
Seminiferous tubules, 948, (fig.) *facing page* 956
Semipermeable membrane, 35
Senescence, 1024
Senility, 1021, 1023
Sensation, 648
  and stimulus, (fig.) 649
  definition, 648
Sense, chemical, 661
  pain, 657
  smell, 661
  taste, 664
  temperature, 652
  touch, 654
Sensory areas of the cerebral cortex, 897, 902
  decussation, 848
  homunculus, 903
  impulses from joints, 769
  paths, (fig.) *facing page* 825
Separated plasma, 577
Serotonin, 400, 946
Sertoli, cells of, 949
Serum, 362
  calcium, 929
  chemistry of, 360
  chorionic gonadotrophin, 980
  iron, 393
  phosphate, 929
Sex chromatin, 1019
  chromosomes, 1015
  cycles of the mouse, bitch and woman, 962
  determination, 1015
  -linked inheritance, 1019
  of cells, 1019
  ratio, 1015
Sexual activity and light, 970

Sham feeding, 242
  rage, 878
Shipwreck, 212
  survivors, 635
Shivering, 200, 206, 212
Shock, 524
  induction coil, 759
  nervous factors, 525
  spinal, 829
  toxic factors, 525
Short sight, 679
Shortening reaction, 844
Silicone and clotting, 364
Simmonds' disease, 1004, 1008
Simultaneous conditioned reflex, 912
  contrast phenomena, 691
Sinu-atrial node, 424, 425, 454
Sinus arrhythmia, 473, 495
Skatole, 281, 349
Skeletal muscle, chemistry of, 772
  structure of, 771
Skin, 651–660
  and heat loss, 200–206
  arteriovenous anastomoses, 206
  circulation in, 519
  colour, 519
  effect of cold, 521
    light, 522
  electrical resistance, 204
  exposure to sunlight, effects, 143, 522
  in old age, 1023
  insulation, 202
  Lewis's H-substance, 521
  pigmentation, 522, 651, 951
  scratch effect: triple response, 520
  sensation, anatomical basis, 659
  senses, modalities, 659
  structure of, 651, (fig.) *facing page* 652
  temperature, 199, 209
  temperature and paralysis, 206
  vasoconstriction, 513
  vessels, mechanical stimulation of, 519
Slater factor, 89
Sleep, 670, 678, 879, 907
  and reticular formation, 908
  electroencephalogram in, (fig.) 907
  temperature in, 208
Sleepy sickness, 909
Small intestine, 251
Smell, 661
Smooth muscle, 801
Snake venom, 364
Sneezing, 564
Snellen's types, 680, 681
Soaps, 20, 267
Sodium, 107
  in cells, 737
  pump, 635
  space, 633
  transfer, 739
Solar radiation, 203
Solution, hypertonic, 37
  hypotonic, 37
  isotonic, 37
  molar, 29
  normal, 29
Somæsthetic area, 903
Somatotrophin, 1005
  and lactation, 993
Sound, localization of, 722
Sound, reinforcement, 723
  shadow, 723
  temporal lobes and, 902
Sounds, breath, 547
  heart, 438
Sour taste, 666
Spastic paralysis, 894
Spatial summation, (fig.) 833
Special senses, 648
Specific dynamic action, 169, 192, 336
  gravity of blood, 375
  irritability, 648
  nerve energy, 648
Specificity of enzymes, 73
Spectroscope, 408
Spectroscopic tests for blood, 407
Spectrum, (fig.) 408, 686
  acoustic, 709
  colours of, 686
Speech, 705
  analysis of speech sounds, 708
  and saliva, 222
  energy involved in, 710
  in cerebellar disease, 875
  laryngectomy and œsophageal speech, 708
  organs of, 705
  development of, 705
  quality, 726
  rate of vibration of the vocal cords, 710
  waveform, 708
Speed of the circulation, 501
Spermatids, 1014
Spermatocytes, 948, 1014
Spermatogenesis, 948, 954
Spermatogonia, 948
Spermatozoa, 949, 968
  metabolism, 950
  viability of, 972
Spherical aberration, 677, 678
Sphincter ani, 277
  ileocolic, 262, 276
  œsophageal, 224, 277
  of Oddi, 252, 257
    urinary bladder, 624
  pupillæ, 901
  pyloric, 231, 234
Sphygmomanometer, 484
Sphingomyelins, 23
Spike potentials, 743
Spinal animal, 832
  cord, 822
    and defæcation, 278
    ascending paths, 825
    Brown-Séquard syndrome, 830
    degeneration in, 825
    descending paths, 825
    hemisection of, 830
    posterior root fibres, 826
    root ganglion, (fig.) *facing page* 824
    section of roots of spinal nerves, effect of, 828
    segment, (fig.) 823
    sensory paths, (fig.) *facing page* 825
    synapses in, 749
    transverse lesions, 815
  ganglion, 734
  nerves, cutaneous areas served by posterior nerve roots, (fig.) 828
  preparation, 832
  shock, 829

Spinal reflexes, 830
  extension, 837
  flexion, 830
  reciprocal innervation, 845
  stretch, 839
Spinocerebellar tracts, 825
Spinothalamic tract, 825
Spiral arteries, 958
Spirometer, 159, 543
Splanchnic nerve, stimulation of, (fig.) 510, 530
  nerves, 261, 510, 942
Spleen, 377, 531
  and blood formation, 386
  and platelets, 400
  blood vessels of, (fig.) 532
    volume, 533
  circulation in, 531
  exercise and the, 532
Splenectomy and platelets, 400
Squint, 679, 700
Stagnant anoxia, 587
Stannius ligature, 427
Stapedius muscle, 711
Stapes, 711
Starch, 15
Starling's law, 446
Starvation, 282, 297, 330
Status lymphaticus, 944
Staub-Traugott phenomenon, 290
Stearic acid, 19
Stercobilin, 281, 415
Stereognosis, 904
Stereopsis, 703
Stereoscope, 702
Stereoscopic vision, 703
Sterility and vitamin E, 147
Sterilization of milk, 175
Sternal puncture, 390
Steroids, 23
  isolated from cortical tissue, (fig.) 937, (fig.) 939
  table, 26
Sterols, 24
  absorption of, 275
Stethograph, 542
Stilbœstrol, 965
Stiles-Crawford effect, 691
Stimulation, faradic, 759
  galvanic, 758
Stillbirths, 1016
Stimulus, definition of, 648
Stokers' cramp, 205
Stomach, 230–250
  absorption in, 268
  and hunger, 214, 232
  arteriovenous anastomoses, 231
  barium meal and, 232
    X-ray appearances after, 232, 233
  digestion, 245
  effect of acetylcholine, 233
  emptying, 234
  fundus, (fig.) 231
  gastric analysis, 246
  hydrochloric acid in, 239
  lymphatics, 231
  movements, 231
  mucosa, 236
  outline, (fig.) 231
  peristalsis, 232
  secretion, 241
Stomach, structure of human, 230
  vomiting, 249
Stratum corneum, 651
  germinativum, 651
  granulosum, 651, 956
  lucidum, 651
Streaming birefringence, 68
Strength-duration curve, 755
Strength of solutions, 29
Streptomyces griseus, 133
Streptomycin and vestibular nucleus, 860
Stress, 939, 942
Stretch receptor, 766, 768
  reflex, 839
Stria vascularis, 714
String galvanometer, 468
Stroke volume, 449
Stromuhr, 501
Stuffiness, 213, 537
Subarachnoid space, 644
Subliminal fringe, 747, 834
Sublingual gland, 218
Submandibular gland, 218
Substantia gelatinosa, 823, 827
  nigra, 851
Substrate, 69
  competition, 134
Subthalamic body of Luys, 885
Successive contrast phenomena, 691
  induction, (fig.) 842
Succinic acid, 75, 85
  dehydrogenase, 75, 84
Succus entericus, 264
Suckling, 993
Sucrase, 265
Sucrose, 13, 14
Sugar transport by placenta, 982
Sugars, 8
  'D' series, 10
  'L' series, 10
Sulphæmoglobin, 409
Sulcus, central, 883
  lateral, 883
Sulphanilamide, 419
Sulphur excretion, 355, 419, 603
  metabolism, 354
  pathways of, (fig.) 354
Sulphydryl (-SH) groups, 74
Summation tones, 712
Sunburn, 522, 657
Sunlight and rickets, 143
Superior cervical ganglion, 746
  quadrigeminal body, 851, 900
Supporting reactions, 846
Suppressor areas, 894
Supraoptic nucleus, 617, 876, 1000
Suprarenal glands, 935
  cortex, 936
  cortex and water balance, 619
  cortical tumour, 941
  extracts, 938
  medulla, 941
Supraoptico-hypophysial tract, 877
Suprarenalectomy, 936
Surface area, 163, 166
    nomogram, 163
  formula for calculating, 167
  height, weight and, of children and adolescents, 167
  tension 256

Swallowing, 225, 564, 712
Sweat, 95, 204
  atropine, 205
  composition of, 204
  glands, 204
  heat loss, 203
    amount of, 203
Sweating, 95
  acetylcholine, 205
  gustatory, 668
  nervous control, 205
  salt loss, 205
Sweet taste, 667
Sweetness, factors determining, 667
Sympathectomy, 805
Sympathetic fibres, (fig.) *facing page* 805
  ganglion, 747
Sympathin, 811
Synapse, 746
  definition, 746
  investigations on superior cervical ganglion, 746
  mid-brain, 751
  refractory period, 747, 751
  spinal cord, 749
Synaptic delay, 751, 832
  knobs, (fig.) *facing page* 745, 746, 824
  potential, 750
  transmitter, 753
Syncope, 523
Syncytiotrophoblast, 976, 977
Synovial fluid, 17
Synthetic œstrogens, 964
Syringomyelia, 658
Systolic blood pressure, 485
  output of heart, 449

Tachypnœa, 555
Tactile sensation, 655
Takayama's test, 407
Target organs, 999
Tartar, 222
Taste, 222, 664, 904
  adaptation in, 668
  bud, 664
  classification, 665
  fibres, pathways of, (fig.) 665
  localization in postcentral gyrus, 904
  nerves of, 664
  temperature of food and, 668
  threshold concentration, 667
Taurine, 355
Taurocholic acid, 355
Tear fluid, 670, 671
Teeth, 101, 104, 107, 140
  ascorbic acid deficiency, 140
  composition of, 104
  fluorine and, 107
  tartar and, 222
  vitamin D deficiency, 145
Telencephalon, 883
  parts of, (table) 883
Teleradiograms, 426
  of chest, (fig.) *facing page* 424
Temperature and thyroid hormone, 201
  after adrenalectomy, 939
  after exercise, 797
  and heart rate, 468
Temperature, balance, 198
  body, 199
  diurnal variation of body, 208
  extremes of body, 209
  gradient, 199
  heat stagnation, 213
  high, effects of, 209
  in exercise, (fig.) 797
  in menstrual cycle, 966, 967, 974
  in sleep, 208
  liver, 200
  low, effects of, 209
  mouth, 200
  neonatal, 208
  nervous control of, 205
  of tissues, 202
  of tissues of forearm, (table) 199
  oral, 207, 208
  rectal, 199, 207, 212
  regulation, 198, 205
  sense, 652
  shivering, 200
  skin, 199, 202, 207, 209
  variations in, 207
  wet bulb, 212
Temporal lobes, 901
  lesions in, 902
Tendon jerk, 749
  organs, Golgi, 771
Tension of gas in water, 571
  of gases in blood, 582
  of muscle, 781
Tensor tympani, muscle, 711
Testa, 179
Terminal buttons, (fig.) *facing page* 745
Test meal, 246
Testis, 948, 968
  and vitamin E, 147
  descent of, 948
  hormones of, 951
  œstrogen production, 953
  pain, 808
  puberty, 954
Testosterone, (formula) 951
  effects of, 950, 954
  in delayed puberty, 954
  propionate in protein anabolism, 954
Tetanus, 782
  of muscle, (fig.) *facing page* 784
  toxin, 790
Tetany, 102, 557, (fig.) *facing page* 929, 929
  and parathyroids, 929
  causes of, 930
Thalamic animal, 862
  syndrome, 881
Thalamus, 879
  connexions, (fig.) *facing page* 881
  dorsomedial nucleus, 881
Thebesian vessels, 426
Theobromine, 59
Thermal insulation, 210
Thiamine, 121, 135, 748, 965
  daily allowance, (table) 190
  deficiency, 122
    heart rate and, 123
    lactic acid and pyruvic acid and, 122
  estimation of, 121
  fat metabolism, 328
  in blood and tissues, 123
    foods, (table) 124

Thiaminepyrophosphate, 122
Thiochrome, 122
Thioctic acid, 124
Thiocyanate space, 634
Thioethanolamine, 62
Thiouracil, 58
Thiourea, 927
Thirst, 216
  and alcohol, 217
    atropine, 216
    blood osmotic pressure, 216
    diabetes mellitus, 308
Thiry-Vella fistula, (fig.) 267
Thoracic duct, 398, 638
  nucleus, 823, 827
Thoracolumbar outflow, 804
Threonine, 41
Threshold, 739, 755, 779
  galvanic, 755
  of audibility, 724
  of feeling, 724
Thrombin, 362, 365
Thrombocytopenia, 368
Thrombokinase, 365
Thromboplastin, 365
  generation test, 366
Thrombosis, 368
Thymine, 57, 65
Thymus, 944
Thyroid gland, 919
    and hæmopoiesis, 395
    and radioactive iodine, 922
    and temperature, 201
    development, 919
    endemic goitre and, 927
    histology, (fig.) *facing page* 924
    hyperthyroidism, 925
    hypothyroidism, 924
    stimulating hormone (TSH), 920
    thyrotrophic hormone and, 920
    thyroxine in, 921
    uptake of ^{131}I, 923
    weight of, 919
Thyroidectomy, 925
Thyrotrophic hormone (TSH), 1002
Thyroxine, 921
Tidal volume, 547
Time, circulation, 504
Tinel's sign, 764
Tissue fluid, 629
  capillary permeability, 636
  extracellular fluid of muscle, 638
  formation of, (fig.) 636
  hydrogen-ion concentration, 640
  lymph nodes, 638
  lymphatics, 638
  relation between intracellular and extracellular fluid, 632
  volume of, 634
Titre, 381
Tm, 599
Tobacco, 405
  mosaic virus, 68
  poisoning, 692
Tocopherol, 146
Tomato, 182
Tone in muscle, 784, 841
  vascular, 514
Tones, difference, 712
  summation, 712
Tongue, 664
Tonometer, 582, (fig.) *facing page* 585
Töpfer's reagent, 247
Total lung capacity, 549
Touch, 654
  localization, 656
  postcentral gyrus and, 902
  projection of, 656
  stability of sensory spots, 655
  threshold pressure, (table) 654
  two-point threshold, 656
Tourniquet, 515, 521
Toxic jaundice, 416
TPN, 62, 127, 316, 317
TPP, thiamine pyrophosphate, 122
Trace, conditioned reflex, 914
Trachea, 537
Tractus solitarius, 664
Training, 801
Transamidination, 352
Transaminase, 129, 341
Transamination, 339, 341
Transferrin, 98, 392
Transfusion, 379, 383
Transients, 708
Transmethylation, 327, 347
Transmitter, adenohypophysis, 1001
Transneuronal degeneration, 824
Transport, iron, 99
Trapezoid body, 722
Traube-Hering waves, 508
Travelling waves, 716, 718, 720
Tremor, 870, 886, 925
  in hyperthyroidism, 926
  intention, 872
Tricarboxylic acid cycle, 316
Trichromacy, 689
Trichromats, anomalous, 692
Triglycerides, 326
3:5:3′-Triiodothyronine (formula), 921
Triolein, 19
Tripalmitin, 19
Triphosphopyridine nucleotide (TPN), 62, 86
Triple intubation, 258
  response, 520
Trisaccharide, 8
Tristearin, 19
Tritanopes, 692
Trophoblast, 976
Tropomyosin, 774
Trousseau's test, 930
True plasma, 577
Trypsin, 73, 76, 252, 265
  chymo-, 253
Trypsinogen, 76, 252
  chymo-, 253
Tryptophan, 129, 185, 349
Tubular excretory capacity, 606
  fluid, 598
  function, 598
  mass, 606
  reabsorption, 598
  secretion, 599
Tubules of kidney, 598
Tunica adventitia, 480
  externa, 956
  interna, 956
  intima, 480
  media, 480
Tryptophan, 185

Turbulence, 500
Twins, 1017
Two-point threshold, (table) 656
Tympanic membrane, 711
Tyramine, 350
Tyrosinase, 77
Tyrosine, 81, 185, 348, 350, 920
Tyrosinosis, 349

Udder, 994
Ultraviolet light and vitamin A, 116
  and vitamin D, 143
  light on skin, 522
Umbilical arteries, 985
  cord, 976
  vein, 985
Uncoupling agents, 319
Undernutrition and pituitary, 954
Units, and measures, 1026
Upper motor neurone paralysis, 829
Unsaponifiable fraction of lipid, 23
Uracil, 57, 63
Uræmia, 622
Urea, 283
  and ornithine-arginine cycle, 342
  clearance, 607
  formation, 342
  protein metabolism and, 342
Ureteric pressure and urinary flow, 597
Ureters, 624
Uric acid, 59, 95, 343, 355
  and purine metabolism, 356
  in birds and reptiles, 357
    blood, 357
Uricase, 69
Uridine, 59
  diphosphate glucose, 63, 296
Urinary chorionic gonadotrophin, 980
  concentration, maximal, 601
  excretion of catechol amines, 943
  excretion of gonadotrophin, 970
  excretion of œstrogens in menstrual cycle, 966
  flow and ureteric pressure, 597
  output and adrenaline, 609
    and emotion, 610, 619
      removal of posterior pituitary, 613
  pH, 621
  sulphates, 355, 419, 603
Urine, acidification, 621
  baby's, 623
  blood plasma and, comparison of composition, 593
  constituents, 603
  colour of, 601
  denervation of kidney and, 592
  diurnal variations in volume, 95, 602
  effect of hypertonic solutions, 615
  extract of posterior pituitary gland and, 613
  ketosteroids, 953
  œstrogens, 953
  pH, 621
  pregnanediol, 953
  protein in, 595
  specific gravity, 603
Urobilin, 415, 601
Urobilinogen, 415
Urochrome, 601
Uroerythrin, 601
Uronic acid, 12
Uropepsinogen, 241
Urticaria, 945
Uterine activity, 959, 963, 964, 975, 990
  muscle, 963
Uterine contractions after birth, 992
  at coitus, 975
  in labour, 990
  in pregnancy, 990
Uterus, 954
  pressure in at parturition, 990
  rabbit and ovarian hormones, (fig.) *facing page* 964
Utricle, 856

Vagal tone, 458
Vagina, 954
  cellular changes in, 961
Vaginal epithelium of rat, (fig.) 960
  smear, 960, (fig.) *facing page* 961
Vagotomy, 259
  and gastric motility, 235
Vagus, 241, 277
  escape, 458
  nerve, 707
  effect on blood pressure, 460
  to heart, 457, 458, (fig.) 460
Valine, 41
  effect of absence from diet of rat, (fig.) *facing page* 185
Valsalva manoeuvre, 436, (fig.) *facing page* 465
Valves in veins, 422
  heart, 425
Van den Bergh test, 416
Van Slyke manometric apparatus, (fig.) 574
Vaporization, latent heat, 203
Variable element, 326
Vascular system, capacity of, 482
  tone, 514
Vas deferens, 950
  ligature of, 951
Vasoconstrictor reflexes, 512
Vasodilatation, antidromic, 511
  of exercise, 516
Vasodilator nerves, 510
  reflexes, 512
Vasomotion, 518
Vasomotor centre, 507
  chemical control of vascular tone, 514
  control, 507
  effect of carbon dioxide, 514
  nerves, 509
  reflexes, 512
    Lovén, 514
    vasoconstrictor, 509, 512
    vasodilator, 512
Vasopressin, 960, 1009
  structure, 1010
Vasopressor effect, 1009
Vaso-vagal attack, 523
Vegetables, 181
Vein and artery compared, (fig.) *facing page* 481
  structure of, (fig.) *facing page* 480
  valve, 422
Veins of arm, (fig.) *facing page* 422
  structure 482

Velocity of blood, 502
  of pulse wave, 493
Venesection and cardiac output, 452
Venom, Russell viper, 364, 368
Venous and cardiac output, (fig.) 452
  blood, gases, 572, 583
  inflow in heart-lung preparation, 442
  pressure, 487, 489
  pulse, 498
  return, 436
  thrombosis, 368
  valves, 481
Ventilation-perfusion relationship, 584
Ventricular pressure, 433, (fig.) *facing page* 457
  extrasystoles, (fig.) 474
Ventriculogram, (fig.) *facing page* 645, 647
Verbal aphasia, 902
Vertigo, 861
Vesicular murmur, 547
  ovarian follicle, 956
Vessels, capacity of, 482
  coronary, 426
  flow in, 506
  nervous regulation of, 507
  pressure in, 482
  structure of blood, 480
Vestibular fibres, connexions of, 858
  nerve, impulses in, 860
  nucleus, 854
  nystagmus, 861
Vibration sense, 828
Villikinin, 263
Villus and absorption, 268
  chorionic, (fig.) *facing page* 976
  duodenal, 260 (fig.) *facing page* 263
  intestinal, (fig.) *facing page* 260, 263
  movements, 263
Virilism, (fig.) *facing page* 940, 941
Viruses, 67
Viscera, afferent impulses, 805
  segmental sensory innervation of, (table) 807
  sensory innervation of, 807
Visceral muscle, 801
  pain, 805
  pleura, 539
Viscerosensory fibres, 805
Viscosity, muscle, 784, 786
Vision, 669
  accommodation, 674
  and cortex, (fig.) *facing page* 897
  binocular, 700
  colour, 689
    blindness, 691, 1019
      theories of, 693
  duplex theory, 685
  field of, 492, 687
  optical properties of the eye, 678
  optics of the eye, 672
  refractive errors, measurement of, 681
  retina, 681
  structure of the eye, 669
  under water, 674
Visual acuity, 685
  pathways, 898, 900
  purple, 686, 698
  purple and anoxæmia, 698
  reflexes, 898
Visuosensory cortex, 888
Vital capacity, 548
Vitamin A, 115–120, 928
  absorption of, 116
  acetate, international standard, 116
  aldehyde, 119
  and night-blindness, 118
  blood plasma, 117
  carotene, daily allowance, (table) 190
  deficiency, 118, 698
  estimation of, 116
  goitre and, 928
  in milk, 120
  potency of foods, (table) 120
  requirement, 120
  ultraviolet light and, 116
Vitamin A_2, 116
Vitamin B complex, 121
  beri-beri and, 123
  deficiency and blood pyruvate, (fig.) 777
  lipoic acid, 124
  nicotinic acid, 127
  nutritional neuropathy, (fig.) *facing page* 128
  pantothenic acid, 129
  pyridoxine, 129
  riboflavin, 125
  thiamine, 121
Vitamin B_{12}, 132, 395
  and pernicious anæmia, 386, 394
  in foods, (table) 133
  tests for, 133
Vitamin C, 136
  and bones, 139
  daily allowance, (table) 190
  in foods, 141
  in tissues, 138
Vitamin D, 100, 142
  daily allowance, (table) 190
  estimation of, 144
  in foods, (table) 146
  international standard, 144
  rickets and, 142
  ultraviolet radiation and, 143
Vitamin D_2 (ergocalciferol), 143
Vitamin D_3 (cholecalciferol), 143
Vitamin E, 146
  deficiency and sterility, 147
  foods, (table) 148
  pregnant rat uterus and, (fig.) *facing page* 148
  reproductive organs and, 147
Vitamin K, 148
  absorption of, 275
  bile salts and, 150, 275
  deficiency and prothrombin, 149
  in various materials, (table) 150
Vitamin K_1 (formula), 148
Vitamin K_2 (formula), 148
Vitamin P, 142
'Vitamines', 109
Vitamins, 109–151
  bacterial synthesis of, 280
  cooking, effect, 183
  deficiency and biochemical lesion, 113
  definition, 109
  liver function and, 418
  list, 112
  minimum requirement, 114
  optimal requirements, 114
  requirements, (table) 190
  saturation and, 113
  transport by placenta, 983

Vitreous humour, 670
Vocal cords, 705
  folds, 705
Vocalis muscle, 705
Volley, 743
Volume and composition of urine, 601
  elasticity curves of aorta, 481
  flow, 501
  mean corpuscular, 387
  of blood, 376
  of plasma, 377
  of red cells, 376
  surface ratio, 762
  units used, 1026
Voluntary apnœa, 557
  hyperpnœa, 557
  muscular contraction, 783
Vomiting, 249
von Frey's bristles, 654
Vowels, 708
  acoustic spectra, 710
  oscillograms of, (fig.) 708

Waking centre, 909
Wallerian degeneration, 763
Warm spots, 652
Water, 93, 629
  balance, 93, 612
  balance and adrenal cortex, 619
  content of body, 631
  content of tissues, 93, 631, 1023
  depletion, (fig.) 635
  distribution in body, 631
  diuresis, 613, 616, 619
  kidney and, 612
  latent heat of vaporization, 203
  loss and thirst, 216
  red blood cell, 94
  requirement and shipwreck survivors, 635
  response to ingestion of water and salt, 613
  retention in pregnancy, 980
  suprarenal cortex and, 619
Waves, Mayer, 508
  Traube-Hering, 508
Wax, ear, 711
Waxes, 21
Weal, 520
Weber-Fechner law, 649, 710
Weber's fraction, 649
Weber's law, 648, 649
Weber's tests, 730
Weigert-Pal method, 824
Weight at birth, 984, 1021
Weight gain in pregnancy, 981
Wet bulb temperature, 212
Wever and Bray phenomenon, 717
Wheat-germ oil, 146
Wheat gluten, 274
  grain, composition of different parts, 178
Wheatstone's stereoscope, (fig.) 702
Whey, 174
Whispering, 706
White cells, 396
  childhood, 402
  enumeration of, 399
  list of, 399
  origin of, 396
White ramus communicans, 805
Wigger's manometer, 435
Withdrawal bleeding, 966
  reflex, 832
Wolffian body, 948
  duct, 948
Work of muscle, 785
  of the heart, 447
  of the kidney, 600
Working, optimum rate of, 784
Wound repair and ascorbic acid, 140

Xanthine, 59, 355
  oxidase, 87, 193, 355
Xanthoproteic reaction in proteins, 55
Xenopus test, 979
Xerophthalmia, 118
Xerostomia, 216

Yeast and thiamine, 124
Yellow enzyme, 87, 125

Zein, 46, 185
Zimmerman reaction, 952
Zinc, 77
  and insulin, 933
Zona fasciculata, 936
  glomerulosa, 936
  reticularis, 936
Zwitterions, 44, 50
Zygote, 1013
Zymogen granules, 223, 224
Zymogens, 76
Zymohexase, 312

O.V.C.N-7/
14/8/89

12.5.97

Acc.No. 3871

Call No. L;3&eGX1 N56

Please return this publication on or before the last DUE DATE stamped below to avoid incurring overdue charges.

Due date	Return Date	Due Date	Return Date
		Reserved.	
20.5.90	8/5/90	1) ~~Miss. Ambili.V.~~	20/5/91
3.6.90	2.6.5		
4.5.91	3.5.91	19.2.93	23.2.93
7.9.91	11/9	17/4/97	10.5.97
~~26/9~~		9.9.98	9/11/98
1.5.93	1/6/93	15.11.01	20/9/07
22.10.93	22.10.93	4/12/01	4/1/2007
6.11.93	29.11.93		
2.4.91			
23.4.91	22.5.91		
7.5.98	2.5.98		

VERIFIED
2013

Textbook of phys.